IS THERE A BEST MODEL?

A RADIOECOLOGICAL CASE STUDY

by

Davide Tarsitano

Thesis submitted to the University of Nottingham
for the degree of Doctor in Philosophy, November 2005

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

OL parameters	Optimised Local parameters
OG parameters	Optimised Global parameters
UA	Uncertainty Analysis
SA	Sensitivity Analysis
AMU	Aggregated Model Uncertainty
NR_p	Normalised Range
MCS	Monte Carlo Sampling
LHS	Latin Hypercube Sampling
OAT	One-at-a-time Sensitivity analysis method
FDA	Finite Difference Approximation Sensitivity analysis method
SRC	Standardized Regression Coefficient
PCC	Partial Correlation Coefficient
RCC	Rank Correlation Coefficient
CoeffVar	Coefficient of Variability
CV	Contribution to Variance
TF	Transfer Factor
CF	Concentration Factor
T_{agg}	Aggregated Transfer Factor
MVO	Multivariable Optimisation
LM algorithm	Levenberg-Marquardt algorithm
MRE	Mean of Relative Error
Ms	Simple model
Mc	Complex model
$Data_{ob}$	Measurements
$Data_{real}$	Real Data
GOF	Goodness of Fit
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion

MIDL	Minimum Description Length
ICOMP	Information-theoretic measure of complexity
CVn	Cross Validation
ML	Maximum Likelihood
D&C matrix	Detail and Complexity Matrix

ABSTRACT

Mathematical models are extensively used to support decision-making in many disciplines. Nevertheless there are not clear standard guidelines to assess models performance. This significantly affects model selection processes, which aim to determine the "best model", among several possible candidates.

Model performance is often measured by the accuracy with which models predictions fit independent observations. However this test assesses only a single aspect of a model. A model selection process should establish the similarities between the constructed and the conceptual model. Therefore it should be based on a comprehensive assessment of the models capabilities, which is the objective of the multi-aspect comparison approach proposed in this work. The innovative aspect of this approach is to create a relationship among four conventional tests, i.e. uncertainty and sensitivity analysis, goodness-of-fit prediction-observations, model complexity and level of details, in order to provide a reliable estimation of the differences between the constructed and conceptual models. Although, model complexity is quantified using a standard approach, a novel methodology is proposed in this thesis, intended to be an intuitive and illustrative approach in creating a linkage between model complexity and level of detail.

Five radioecological models have been considered: SAVE rural model, TEMAS rural model, SAVE semi-natural model, FORM and RIFE1. The results show that there is a limited resemblance between these models and the respective conceptual models. This is due to low prediction accuracy (RIFE1 and FORM); high level of uncertainty (SAVE rural); sensitivity to parameters which is not consistent with the current understanding of radiocaesium behaviour in the environment (TEMAS and SAVE rural).

The SAVE rural model has been revisited in order to increase the similarity between the constructed and conceptual model. The resulting model prediction shows lower degree of uncertainty and there is a significant agreement between the model sensitivity results and the general understanding of the processes affecting Cs soil-to-plant transfer. Nonetheless the revised model does not show higher prediction accuracy than the original model.

It is concluded that a reliable methodology for model selection should be based on a comprehensive investigation of each considered model aspect and that there is

not a single best approach. The methodology proposed in this work has been successful in the case of the five radioecological models studied.

Acknowledgments

The first person I would like to thank is my supervisor, Prof. Neil Crout, for his support guidance and patience throughout this project

I would also like to thank all the people from the Environmental Science group at University of Nottingham; a special thought is for: Glen “Black Sabbath are melodic”, Marcello “March”, Tolis “I have a good deal”, Melissa “Mel”, Tim, Lee “do I look gay in this”, Vicky, Bryan, Ed, John and Darren to have made these four years of work a good laugh.

I would also thank the “10 College Road” group, in particular Gwydion “you did not manage to convince me that Cider is of any good”, Taki and Satoru for the innumerable nights spent at the Bell View and Greek Taverna speaking about ….. nonsense. In addition I am really grateful to Ian Kelso to have bet on me and to have given me “the opportunity”.

I would like also to mention all my friends in Nottingham, in particular Alex “Babra&Adamus”, Anne “technically I won our last game”, Arthur “my negative”, Marcella, Monica, Nicy “Miss Hoegaarden”, Nick, Olga “Mr Green”, Oliver “Mr Bond”, Pericle, Richard “Ricky Cunningham”, Rolf “the Flemish”, Joe “the Bond girl”, Sam, Margherita “Ebbuggia!!!” and all the others I might have forgotten. In addition the Canal House to have adopted us in our long nights seeking for “a later last order”.

Special thanks to the Italian mob (AV): Alessandro “Fiorello”, Flavio “Pino, Vuoi balare – Con chi? – Con te!!”, Lele “Frusta”, Marco “Trota, che te serve – to trovo”, Roberto “Bruco” for our unfailing early breakfast (i.e. 12:00) at “Praticello” and for having been so perfect in “cazzeggiare”.

I would also like to thank my family to have been of an incredible support and fun to be with every time they came for the exploration of the “Grigia Albione”, my dad “er barone”, my mum “a baronessa”, Gulia “The snooker girl”, Zio Fabrizio, Zia Elena and my two little cousins Elisa e Francesca.

Last, but by no means least, I would like to thank Julie who has always been there when I needed her help and for being so patient with my long moments of memory failure. I would like to thank her because the thesis would not have been done without her precious support.

DEDICATION

To Orazio and Anna for everything they have done for me

1. INTRODUCTION

Mathematical models are ubiquitous in many disciplines, from the medical to the environmental sciences. Nowadays, mathematical models are in the public eye, due to their implementation, sometimes controversial (Summerhayes, 2005; Basgall, 2005), to predict changes in the climate as consequences of the increase of greenhouse gases into the atmosphere (McCarthy, 2005; Colacino, 2005).

Generally, modelling is undertaken for two main reasons and firstly to support decision-making. In this case the model can be described as "purely predictive", as the interest is to establish the behaviour of the system under investigation in particular conditions, when it is impossible or undesirable to experiment on the system itself. Relevant examples are the radioecological models developed to establish the level of contamination of humans and foodstuffs in the event of radionuclide deposition (e.g. SAVE (Howard et al, 1999); MOIRA (Monte et al, 1999); ECOSYS-87 (Muller and Pröhl, 1993); PATHWAY (Whicker and Kirchner, 1987); RODOS (Ehrhardt et al, 1993)). The second reason, which may require the application of modelling, is in research projects where the objective is to better understand a system under observation. In these cases the model has a more descriptive role as it summarises the current knowledge on the considered system (e.g. RIFE1).

Often models initially developed as part of a scientific project are later successfully implemented for decision support. Nonetheless, these applications have different aims. A model, which supports decision-making, focuses on producing a model which could be applied to a wide range of scenarios. This implies, for example, that the input parameters should refer to measurements which are easily accessible for any user. For instance the model proposed by Absalom et al (2001) predicts the radiocaesium activity concentration in plants as a function of pH, exchangeable potassium, clay and organic matter content. These are soil characteristics commonly measured. On the other hand, a scientific model is developed to test the current understanding of the system behaviour. Therefore it is not subject to comparable restrictions on the type and quantity of inputs used. For instance, RIFE1 has been developed to study the behaviour of radiocaesium in forests and it requires the tree root distribution, described as the percentage of roots present in each soil layers, as input parameters, which can only be measured as part

of a comprehensive fieldwork analysis. This makes the model application, as decision support, more problematic.

In addition, models may differ significantly in their structural design. Generally, two categories can be identified: mechanistic and empirical models. Mechanistic models aim to simulate several processes which take place in the system under investigation. Therefore, the estimation of the final output is performed using "mid-process steps", which could be regarded as sub-models.

On the other hand, empirical models are usually statistically based and are developed to describe the relationship between inputs and outputs without considering in detail the processes that govern the system behaviour.

Mathematical models can differ significantly due to their outcome or structure. Consequently, it is essential to have a full overview of a model capability and performance before it is used as a support tool in decision-making. Nevertheless there is no general methodology which can be adopted for a comprehensive analysis of model performance. This has evident effects on model selection processes where the "best model" must be selected from several possible ones, for application to decision making.

Generally, the prediction power is the main model characteristic that is used to establish which model should be selected. The predictions are compared with independent measurements and the model that shows a higher level of accuracy in predicting the observation trend is regarded as the best model and therefore selected. Avila et al, (2001) and Goor and Avila (2003) undertook a model comparison exercise, where models have been mostly assessed on the basis of the prediction accuracy (IAEA, 1996; IAEA, 2002). Although the goodness of fit of the model predictions is an important measure of model performance, model selection process should not be solely based on this aspect. The selected model should be the one that provides the best fit to new datasets. If it is able to adequately simulate the system behaviour, it can be regarded as having a high level of generalizability, which is the capability of a model to produce an accurate prediction using a dataset not used for the parameterisation.

However the connection between goodness-of-fit and model generalizability is not straightforward. This work will illustrate how the model generalizability is related to the complexity of the model and that the goodness-of-fit is only the end result of the complexity-generalizability relationship. In addition, the test of

goodness-of-fit may not be a true assessment of model generalizability as independent datasets may not be easily available. For instance, the large majority of field data available to develop radioecological models are measurements carried out following the Chernobyl accident in 1986. Consequently, there is the risk that the high degree of accuracy showed by a model is not due to prediction power but to the resemblance between the observations used for the test and for model development.

Another approach for model selection, primarily adopted for statistical models, is the use of several statistical indices (e.g. AIC and BIC (Myung, 2000)). In this case, model selection is based on the goodness-of-fit and the level of model complexity.

Model complexity, generally quantified as the number of adjustable parameters, has been demonstrated to have a strong impact on the model generalizability (e.g. Myung, 2000; Pitt and Myung, 2002). A more in-depth description of the implementation of these indices is presented in chapter 5.

However, the use of either a goodness-of-fit test or even the statistical model selection indices provides only a partial description of the real model performance. They do not provide sufficient information to establish the resemblance between the conceptual and constructed models.

The concept of conceptual and constructed model should be a key element in any model assessment exercise. The conceptual model is the ideal outcome of the modelling project. In other words it is what modellers aim to produce. Its response to variation of input parameters accurately reflects the knowledge of the system studied. The uncertainty on the predictions is as low as achievable, assuming that the system understanding is correct, and that it has a high level of generalizability, which is the ability of a model to accurately describe the trend of an independent dataset.

On the other hand the constructed model represents the outcome of the modelling project. The constructed model can sometimes have little resemblance with the conceptual model for several reasons. For instance the inadequate mathematical transposition of the conceptual model relationships into the constructed model, with the results that key processes in the conceptual model have little effect on the variance of the constructed model prediction.

A model selection process should be able to establish the degree of similarities between the conceptual and constructed model. However, the literature does not present comprehensive methodologies, which allow the differences between the

conceptual and constructed model to be established. Brooks and Tobias (1996) have suggested guidelines to estimate the model performance considering several model characteristics. However, the authors provide only a general overview of the tests that could be adopted without any practical application.

The work described in this thesis aimed to compare five radioecological models, i.e. SAVE rural, TEMAS rural, SAVE semi-natural, FORM (TEMAS semi-natural) and RIFE1. In order to accomplish this an integrated model evaluation approach is developed based on a number of established methods.

1.1 MODEL COMPARISON METHODOLOGY

The "multi-aspects" comparison approach proposed in this work is not based on novel assessment methodologies, as it comprises conventional tests, i.e. uncertainty, sensitivity analysis, goodness-of-fit predictions-observations, model complexity and level of details. The innovative aspect of this work is to create a relationship among these methods and to provide a comprehensive assessment of models performance, which aims to establish the level of resemblance between the constructed and the conceptual model. Consequently, it could be regarded as an application of the guidelines proposed by Brooks and Tobias (1996).

The proposed multi-aspects comparison approach is divided into five performance elements:

1. The accuracy with which the model fits historical data.
2. The level of resemblance between the model sensitivity to parameters and the current knowledge of the system dynamics.
3. The level of uncertainty of the model outputs.
4. Process resolution; does the model present an adequate level of detail considering the aim for which it has been developed?
5. Model complexity. The model should have the lowest achievable complexity level. This could be quantified as the number of parameters which participate in the model parameterisation

Models are implemented to provide accurate predictions, therefore the degree of agreement between the model outputs and a set of independent observations is an essential part of this methodology (performance element 1). Nevertheless, in order to establish the accuracy of model predictions it is essential to account for the uncertainty associated with the model outputs. This can be determined performing an uncertainty analysis on the model under investigation (performance element 2).

The analysis of the model sensitivity establishes the fractional contribution of model factors to the uncertainty of the model predictions (performance element 3). In addition the results obtained from the sensitivity analysis can be used to address the following points:

(a) Does the model resemble the system investigated?
(b) Which factors contribute the most to the output variation and may therefore require additional research?
(c) Which model parameters or variables can be considered insignificant, in the sense of not affecting the variation of the output and can therefore be potentially eliminated from the final model?

The model applicability and generalizability are strongly affected by two model characteristics: level of detail and complexity (performance elements 4 and 5 respectively). The former affects the model applicability; as each model detail is a potential output since it simulates system processes. Consequently, a detailed model can be applied more widely due to its number of outputs. On the other hand model complexity, often quantified as the number of adjustable parameters, has an effect on the generalizability. In general terms model complexity and goodness-of-fit are positively related, the higher the complexity, the more accurately the model is able to fit a wider range of observations. However, if the level of complexity is too high, the generalizability is reduced (Pitt and Myung, 2002).

In addition to these performance elements, the selection of the best model should account for other model characteristics, i.e. quality of inputs, outputs and software support material, which can improve the model applicability.

This work is described in the subsequent six chapters of this thesis, outlined below:

Chapter 2 – *Aim: Provide the reader with background information on the models used.*

The chapter gives a detailed overview of the five radioecological models that are investigated.

Chapter 3 – *Aim: Estimate the prediction uncertainty and the model sensitivity to parameters for the radioecological models.*

The chapter is divided into two parts: firstly an overview of different techniques to perform an analysis of uncertainty and sensitivity is given. Secondly, the five radioecological models are investigated for uncertainty and sensitivity and the results are discussed.

Chapter 4 – *Aim: Improve model performance based on the results of the uncertainty and sensitivity analysis.*

The model developed by Absalom et al (2001) has been revised based on the results obtained from the uncertainty and sensitivity analysis described in chapter 3.

Chapter 5 – *Aim: Test the prediction accuracy using two independent datasets.*

The accuracy of the models to predict Cs activity concentration in several foodstuffs for rural and semi-natural scenarios is tested. The results are discussed and the model performance compared.

Chapter 6 – *Aim: Provide an overview of the complexity and level of details for the considered models.*

This chapter provides firstly an overview of the concepts of model complexity and level of detail. The models are then compared on the basis of their level of detail and complexity, implementing a new approach which has been developed in this work.

Chapter 7 – The final chapter is devoted to summarising the model performance of each investigated model and provides a discussion on the level of resemblance between conceptual and constructed model for each model studied.

2. MODELS OF RADIONUCLIDE TRANSFER IN THE HUMAN FOOD CHAIN

Radionuclides released into the environment can pose a potential hazard for human health if they enter the food chain. Therefore the assessment of radionuclide transfers from soil to plants and thereafter to animals is a key aspect of radioprotection.

Radionuclides (e.g. radon gas) are commonly found in the environment; nevertheless the hazard for humans largely comes from anthropogenic sources. The intensity of radionuclide release varies enormously from source to source. The most noticeable releases, in terms of level of contamination and area affected, are the atmospheric weapons testing programs of USA, Britain, China, France and USSR carried out between the 1950s and 1980s; the accident at the military facility at Windscale (UK) in 1957 and at Chernobyl (Ukraine) in 1986 (UNSCEAR, 1993). The latter is regarded as the worst nuclear accident in history, due to the scale of the contamination (Smith and Beresford, 2005).

The Chernobyl accident is particularly important in radioecology and radioprotection as it has enabled scientists to investigate the effects of radionuclide release, in the short and in the long term, on natural (e.g. Nimis, 1996; Skuterud et al, 2005), rural (e.g. Ehlken and Kirchner, 2002; Zhu and Smolders, 2000; Albers et al, 2000) and aquatic ecosystems (e.g. Davison et al, 1993) due to the widespread deposition. A significant part of these studies focused on human exposure and on developing countermeasure strategies to limit the propagation of the contaminants through the food chain, as this is the main pathway that poses a threat to human health.

This chapter has two objectives: firstly to provide the reader with a general understanding of the major exposure pathways for humans following a radionuclides release in the environment, and secondly to present a general overview of the radioecological models investigated in this work. For more detailed accounts of these models the reader will be referred to appropriate source material.

2.1 OVERVIEW OF HUMAN EXPOSURE PATHWAYS FOR RADIONUCLIDES AND COMMON MODELING APPROACHES

Following the release of radionuclide into the environment, humans are exposed to radionuclide contamination through four main pathways (Figure 2.1):

- Inhalation of radioactive particles during the deposition period. This can be particularly significant in the short term.
- Ingestion of contaminated food, which can be regarded as the main source of human dose in the long term.
- Ingestion and inhalation of water. This aspect has generally a minor contribution to the human contamination due to the rapid decline of radionuclide activity concentration in lakes and reservoirs (Smith et al, 2005b).
- External exposure from contaminated surfaces.

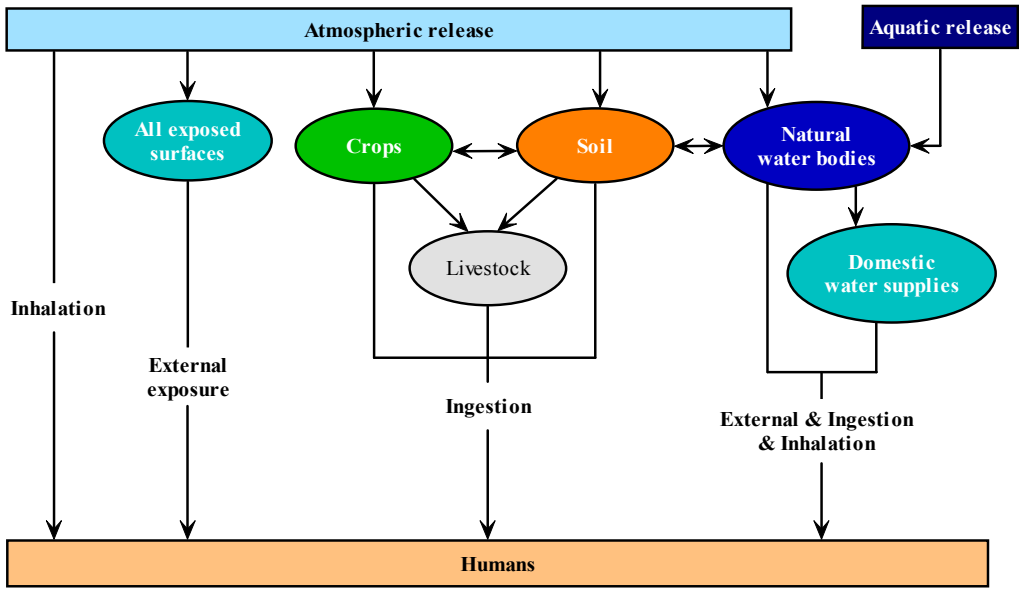


Figure 2.1 Pathways for radionuclides to reach humans. The models investigated concentrate the attention on the fluxes between atmospheric deposition-crops-soil-livestock and humans (Cox, 2004).

The radioecological models considered in this work have been developed to simulate radionuclide behaviour in terrestrial environments, i.e. rural and semi-natural environments. Terrestrial pathways are discussed in more detail later in this

chapter. With regards to aquatic pathways, the work of Monte (Monte et al 2003; Monte et al 1994) and Smith (Smith et al, 2005b; Smith et al, 2000) are suggested.

Radiation exposure can vary considerably over time and spatially, resulting in different levels of dose and pathways of human contamination. This is due to the differences in radionuclide half-life and behaviour in the environment. For instance I^{131} and I^{133} can account for a large proportion of population dose. However with a half life of 8 days and 21 hours respectively, they solely pose a threat in the short term after deposition. In the UK the estimated dose equivalent to thyroid from I^{131} ranges between 590 to 1600 and 84 to 220 μSv for infants and adults respectively. (UNSCEAR, 1988). On the other hand, the fractional contribution of radiocaesium external and internal dose in the overall population dose is strongly related to the type of soil considered (Table 2.1). On a black earth soil, which has the strong ability to fix radiocaesium due to its high clay content, the radiocaesium is not bioavailable for plant uptake and therefore the external dose accounts for more than 60% of the overall population dose under a typical scenario (Smith and Beresford, 2005). In contrast, on a turf-podzol soil, where the transfer of radiocaesium from soil to grass and therefore animals is high, the internal dose is a key factor in the population dose for the same scenario (Smith and Beresford, 2005).

	External dose (%)	Internal dose (%)
Black earth soil	67	33
Turf-podzol (sandy) soil	26	74

Table 2.1 Contrasting contributions of internal and external dose rates to the overall dose in two different soil types (Smith and Beresford, 2005).

Each radionuclide is therefore characterised by a particular behaviour and pathway to enter the food chain. This current work will only concentrate on radiocaesium since it is the principal focus of the models studied.

After Chernobyl, the scientific community investigated in great detail the behaviour of radiocaesium in the environment. This was primarily due to the large extent of the Cs deposition in Europe in comparison to other radionuclides (De Cort, 2001). In addition radiocaesium was of a particular concern due to its bioavailability.

The chemical characteristics of radiocaesium are similar to those of the potassium, which is an essential plant nutrient. Cs is very soluble, consequently in

the scenario of wet deposition or rainfall following deposition, radiocaesium migration through the soil profile may be facilitated (Nimis, 1996). In general terms, Cs is strongly absorbed by clay minerals. Typically in soils having high clay content, less than 10% of the radiocaesium is in the exchangeable form (Smith et al, 2005a). Several studies (e.g. Valcke and Cremers, 1994) have demonstrated that radiocaesium remains available for uptake by plants, whereas it becomes fixed (i.e. unavailable for plant uptake) on clay minerals.

Although radiocaesium is not an essential element for plants, it is absorbed by roots using the same mechanism as potassium. Therefore high transfer of radiocaesium from soil to plant is expected in mineral soil with low nutrient content. Detailed accounts of the factors that control radiocaesium uptake can be found in the literature, e.g. Frissel et al (2002), Ehlken and Kirchner (2002), Facchinelli et al (2001).

In an agricultural scenario, crops are mainly contaminated through root uptake. Foliar absorption represents an important process only in the short-term after deposition. Radiocaesium absorbed by roots is subsequently translocated to other parts of the plant. Animal contamination occurs through ingestion of vegetation and soil (Figure 2.2). Consequently radiocaesium present in crops and rural food products (e.g. dairy milk) is the source of internal dose to humans.

In forest ecosystems a large percentage of the fallout is intercepted by the tree canopy. Melin et al (1994) estimated that in an evergreen spruce stand, 90% of the deposition is intercepted by the canopy, while deciduous stands of beech or birch intercept less than 35%. In addition the time of the year when the deposition occurs may play an important role in the fraction of radionuclide deposited on canopy and soil.

Figure 2.3 shows schematically the fluxes of radionuclides in a forest. Several processes, as leaf weathering and re-deposition on soil, are similar to those considered for rural environments. A detailed overview of radiocaesium behaviour in forests in the short term after deposition can be found in the literature (e.g. Nimis, 1994).

Important forest pathways for human contamination are understorey vegetation, mushrooms, which can accumulate significant radionuclide activity concentration (Skuterud et al, 1997; Kammerer et al, 1994), and game (Ahman et al 2001; Skuterud et al 2005; Zibold et al 2001). In addition to the internal dose exposure, forests can

influence the external dose to workers in the timber industry, i.e. occupational exposure, and general population as they are often used for recreation purposes. Furthermore, forests are used for timber production and the radiocaesium activity concentration in timber can represent an exposure pathway for humans (Gommers et al 2005).

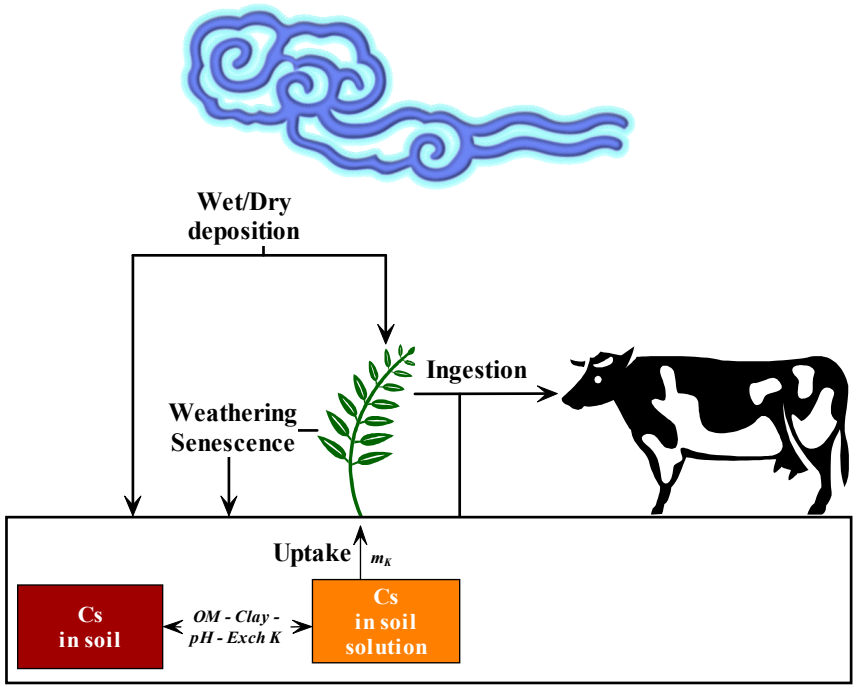


Figure 2.2 Schematic representation of the rural pathways

The migration of radionuclides through the food chain takes place in two phases: firstly the uptake of the contaminant from soil by plant roots and secondly the ingestion of those plants directly by humans or by animals, leading to the contamination of food products (e.g. milk and meat). Therefore the soil-to-plant transfer of radionuclides is one of the most important pathways leading to human exposure.

The estimation of radionuclide transfer from soil-to-plant and subsequently from plant-to-animal is therefore an important aspect of any radiological assessment model. The radionuclide activity concentration in plants is traditionally estimated using aggregated transfer factors and transfer factors (Gerzabek et al, 2003).

The aggregated transfer factors (T_{agg}; equation 2.1) are defined as the ratio of the activity concentration in plant (Bq/kg fresh weight or Bq/kg dry weight) and the total deposition on soil (Bq/m^2). The concept of aggregated transfer factor has been developed in order to overcome the difficulties arising from defining the distribution

of radionuclides concentration in the multilayer soil structure. Nevertheless the T_{agg} estimation of radionuclide activity concentration in vegetation presents several limitations. Firstly, this approach provides only a static snapshot of plant contamination. Secondly it does not take into consideration the effect of root distribution in soil profile or the effects of soil characteristics, e.g. pH, on the radionuclide bioavailability. However, the T_{agg} is a useful approach that can provide a quick estimation of the level of contamination of the plant considered.

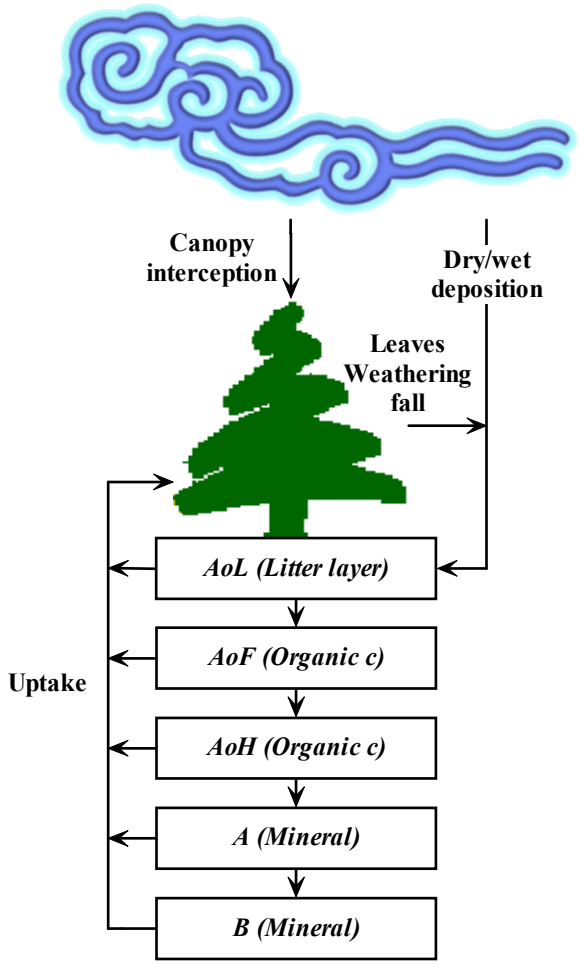


Figure 2.3 Schematic representation of the radiocaesium fluxes in forests (Shaw et al, 2002).

$$T_{agg} = \frac{Plants\ (Bq/kg\ fresh\ or\ dry\ weight)}{Soil\ inventory\ (Bq/m^2)} \qquad (2.1)$$

Transfer factors (*TF*; equation 2.2) referring to a standardised soil depth are defined as the ratio between the activity concentration in plant (Bq/kg fresh weight or

Bq/kg dry weight) and the activity concentration in the considered soil layer (Bq/m^2). This approach is based on the hypothesis that the radionuclide contamination is homogenously distributed within the rooting depth of agricultural plant. This may be proved in the case of agricultural soil where the topsoil layer structure is rather simple. However this approach may have a limited application with soil having a more complex multi-layer structure. In this scenario radionuclides may not have a uniform vertical activity concentration distribution.

$$TF = \frac{Plant\ (Bq/kg\ wet\ and\ dry\ weight)}{Soil\ (Bq/m^3)} \tag{2.2}$$

The literature reports several sets of transfer parameters and their variability can be more than two orders of magnitude (e.g. IAEA, 1994; Frissel et al, 2002; Nisbet and Woodman, 2000). It is evident that they are site and species specific. This variation is due to several factors, which can be divided into two categories: firstly experimental design and secondly the lack, for a long time, of a standard in experimental protocol for the estimation of the TFs or T_{agg}s. The activity concentration in plant was estimated considering only the edible or the inedible part of the plant or in some case the whole plant. In addition, the measurement of the weight varied, dry or fresh weight. In some cases, the transfer factor was considered in terms of unwashed plant, consequently soil particles present on the leaf surface were included in the measure. In other studies the plant was washed, removing soil particles. On the other hand, the uncertainty in soil measurements is due on to variety of soil depth considered, which affects the estimation of the radionuclide activity concentration. Finally the plant morphology may vary if the plants are grown in controlled conditions, e.g. greenhouses. This may affect their nutrient uptake and their development. This lack of standards forced the International Union of Radioecology (IUR) to set measurement protocols (IAEA, 1994). The second category that can affect the variability of TF is the soil characteristics, e.g. pH, nutrients and clay content.

Therefore a standardised method to measure radionuclide activity concentration in plant and soil does not prevent the TF and T_{agg} approach to be characterised by a degree of uncertainty due to the variability of the scenario characteristics.

Consequently, it is difficult to obtain a general TF value as different species react differently in similar environmental conditions. In addition plants belonging to the same species may have different levels of contamination due to the soil characteristics.

The transfer from plants to animals is commonly described using a transfer factor (F_m and F_f for milk and other animal products respectively) and it is defined as the amount of animal's daily intake of a radionuclide that is transferred to one kilogram of animal product at equilibrium (IAEA, 1994). These TFs are also characterised by a degree of variability that may limit their implementation, although this is not as severe as in the soil-plant case.

2.2 MODEL DESCRIPTIONS

The radioecological models considered have been developed to predict the Cs activity concentration in ecosystem compartments in order to establish the potential harmful effects on humans due to the consumption of food products. An overview of the five models is presented below.

2.2.1 SAVE DSS

The objective of SAVE project was to develop a computerised system which was able to estimate the spatial and temporal variation of radiocaesium in rural and semi-natural environments.

SAVE DSS can be divided into three sub-models (Figure 2.4): rural, semi-natural and countermeasures modules. For the purpose of this work the countermeasure model is not described as the focus is on the rural and semi-natural models. The rural model is a semi-mechanistic soil-plant model developed by Absalom et al (2001) which estimates the Cs activity concentration in rural food products, e.g. dairy milk and beef. The SAVE semi-natural model has been developed within the SAVE project to provide predictions ot the level of contamination of semi-natural products (e.g. mushroom and game meat).

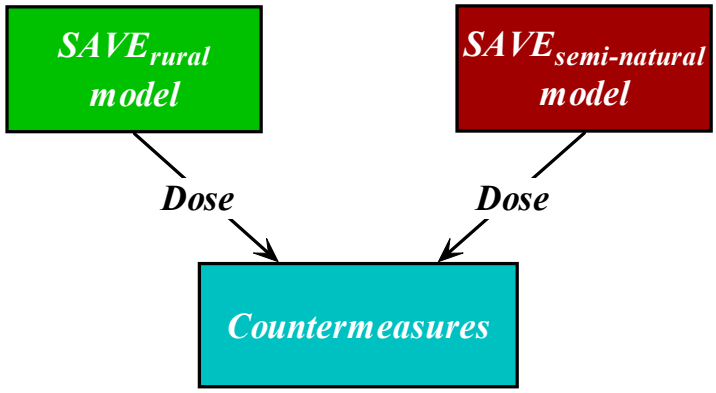


Figure 2.4 SAVE DSS structure. The system is divided into three modules: the rural model (to estimate the Cs activity concentration in rural food products), the semi-natural (to calculate the Cs activity concentration in semi-natural products) and finally the countermeasure module (to estimate the countermeasure effectiveness in terms of dose reduction.

2.2.1.1 SAVE rural model

The SAVE rural model is based on the semi-mechanistic model developed by Absalom et al (2001), which relates the radiocaesium bioavailability to soil characteristics such as clay, organic matter, pH and exchangeable K^+ (Figure 2.5).

Figure 2.5 The Cs availability for plants uptake is determined by five soil characteristics. The Cs solution-absorption equilibrium is significantly affected by organic matter and clay content, pH and potassium in solution (m_k), a limited influence is determined by the NH_4^+. On the other hand the total Cs absorbed by plants is primarily affected by the potassium in solution.

Soil solution potassium is the main competitor of Cs for plant uptake. The model estimates the soil solution K^+ concentration as function of the exchangeable K^+ and the distribution between the organic and inorganic cation exchange capacity is estimated using clay, organic matter content and soil pH (Figure 2.6).

The distribution of Cs between the sorbed and solution phase is calculated using the distribution coefficient (k_d), which is considered to be a function of the specific (clay) and non-specific (humus) K^+ competition coefficients and the potassium concentration in soil solution (Figure 2.6).

The radiocaesium concentration factor (CF), which represents the ratio between the radiocaesium activity concentration in plant and the concentration in soil

solution, is estimated as function of the potassium in soil solution (m_k), using a relationship described by Smolders et al (1999).

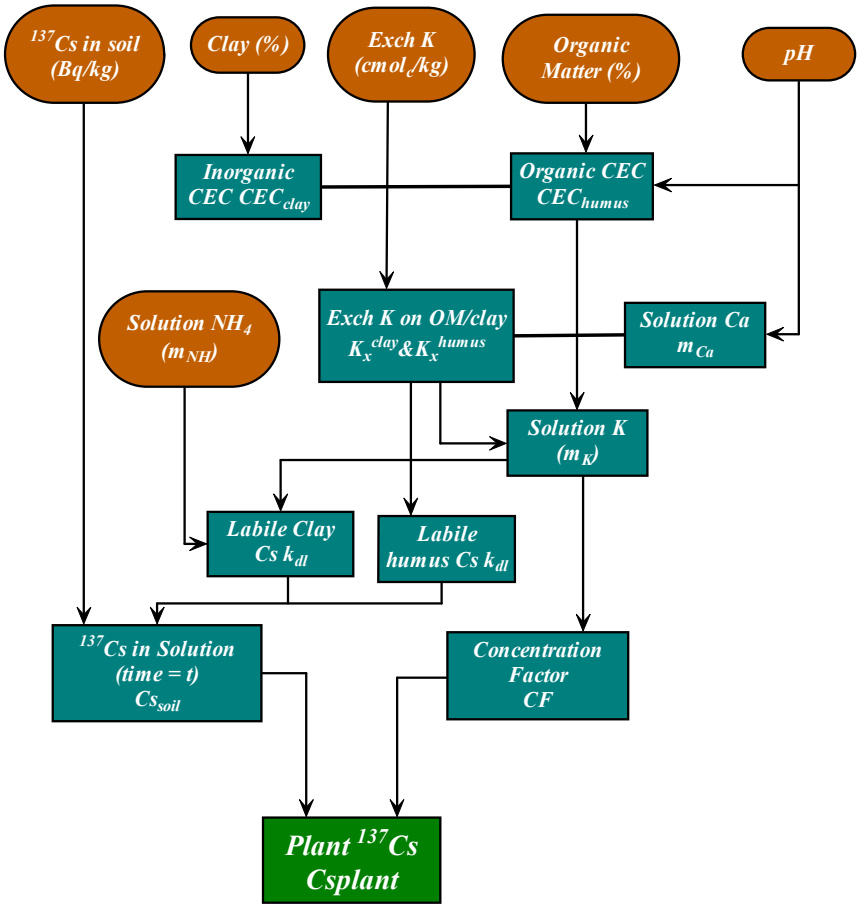


Figure 2.6 Schematic representation of the soil-plant transfer model. Model inputs (oval, brown) and outputs (square, green) are reported (Absolom et al, 2001).

Cs bioavailability is also time dependent due to processes of fixation primarily by clay minerals and leaching from the root zone. In order to simulate the time dependency a reduction factor (D) ranging between 0 and 1 is used. D is calculated using a double exponential equation. The Cs fixation is assumed to be primarily affected by clay minerals, therefore D is adjusted to account for the partitioning of radiocaesium between sites on clay minerals and organic matter.

The model parameterisation was performed using observations from Sanchez et al. (1999), which comprise measurements of predominantly organic rich soil, and data from the Smolders et al (1997) study, from mineral soils.

The Cs activity concentration in animal food products is calculated by using the product between the equilibrium transfer coefficient (F_f and F_m d/kg and d/l,

respectively for milk and meat) and the animal daily dry matter intake (kg/d) and the Cs activity concentration in the feeds (Bq/kg).

A more detailed model description is provided in chapter 4, where an alternative formulation of this model is proposed.

2.2.1.2 SAVE DSS semi-natural model

The model implemented in SAVE DSS used to estimate the radionuclide activity concentration in semi-natural food products uses a simpler approach than the SAVE rural model, namely is the implementation of T_{agg}.

The T_{agg} allows the calculation of the product contamination by simply using the deposition (Bq/m^2). For some food products, i.e. mushrooms, where observations were available the T_{agg} values are adjusted through the use of a radioecological half life, in order to introduce a time dependency into the radionuclide availability.

2.2.2 TEMAS DSS

The main objective of the TEMAS DSS was to provide a management tool to assist decision making for restoration strategies in the event of a radionuclide deposition (Montero et al, 2001). The system has been designed to be applied on a local scale i.e. farm. The terrestrial environments that are considered are: rural, semi-natural and urban (Figure 2.7). However, the urban model has not been considered in this work due to lack of independent observations.

2.2.2.1 TEMAS DSS rural model

The TEMAS rural model is an empirical model which estimates the radiocaesium and radiostrontium activity concentration in plants as a function of physical and chemical soil characteristics, i.e. soil density and exchangeable K^+. (Montero et al, 2001)

The objective of the TEMAS project was to develop a rural model that could be successfully applied in areas where a limited amount of information on the considered scenario is available. In order to achieve this the TEMAS rural model is supported by a database where most of the model inputs are stored.

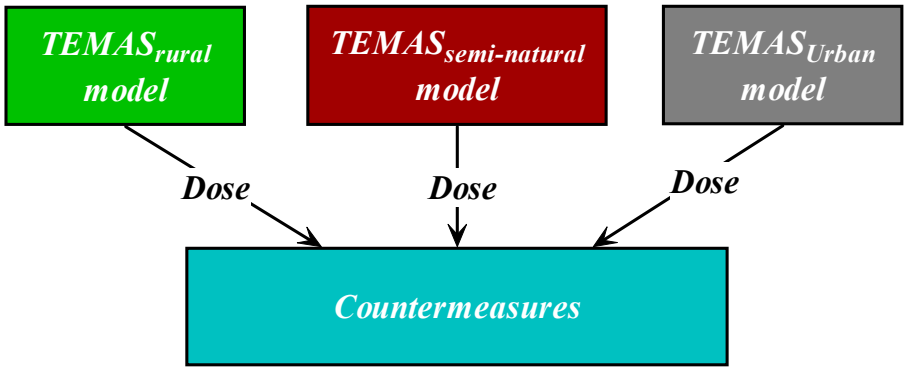


Figure 2.7 The TEMAS DSS is formed by four modules: TEMASrural model, TEMASseminatutal model (FORM), TEMASurban model and finally the countermeasure module which assesses the best countermeasure based on the economic and dose reduction effects.

It is essential to distinguish between specific inputs (I_s) and generic inputs (I_g). The former refers to measurements that users are required to introduce in order to run the model. They are consequently specific to the scenario considered.

Three specific inputs are required and (Figure 2.8):

- exchangeable potassium (for radiocaesium prediction)
- exchangeable calcium (for radiostrontium prediction),
- percentage of organic matter, clay and sand.

Three sets of data are included in the database. Each set refers to a potential scenario and is associated with a set of soil characteristics, which are used by the model to estimate the contamination of the edible and inedible parts of the plants.

The measurements of percentage organic matter, clay and sand are primarily used to classify the scenario under investigation. Therefore, using a decision tree based on these observations, the most appropriate soil scenario is selected and the soil characteristics are implemented in the model evaluation.

The Cs or Sr activity concentration in the edible and inedible part of the plant is calculated as the product of a TF value and the radionuclide concentration in soil solution. However, TEMAS rural does not use a single TF value for each considered crop. Instead, each TF value is described by a uniform distribution, where the geometric mean of the best estimated value reported in the literature has been considered to be a good representation of the minimum transfer while the upper 95% confidence limit is assumed to represent the maximum transfer.

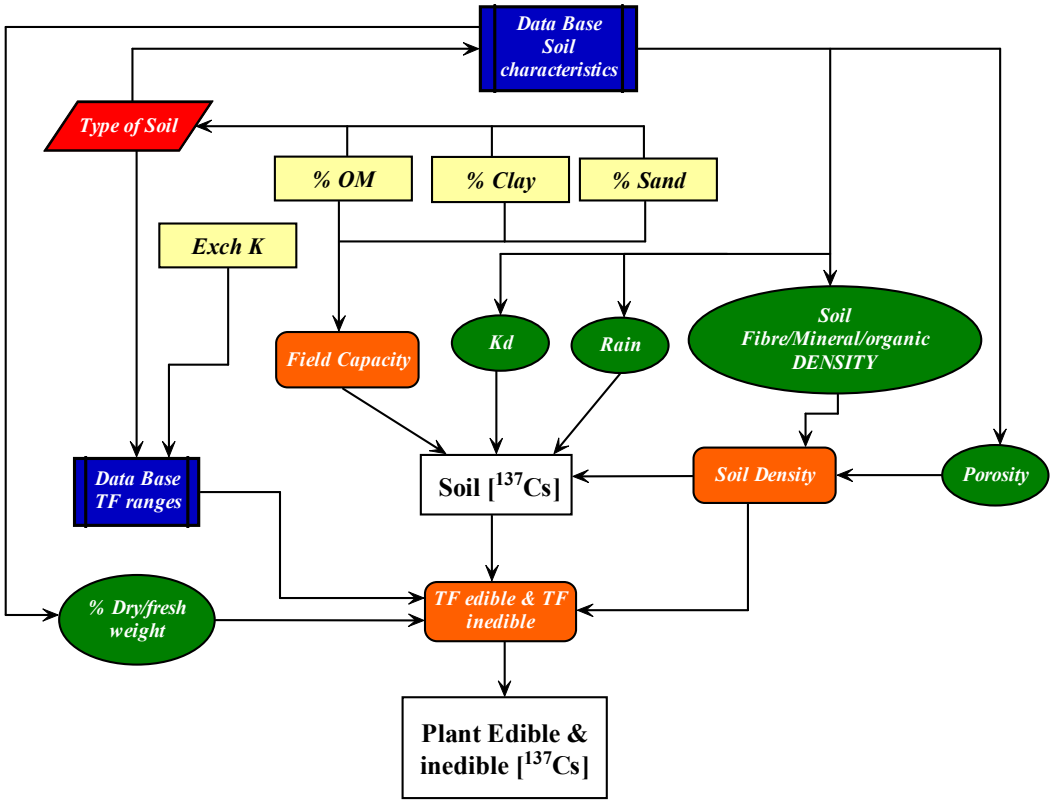


Figure 2.8 Schematic representation of TEMAS rural model. The model outputs (white square), inputs included in the database (green circle), processes estimated by the model (orange square) and model input (yellow square) are indicated.

2.2.2.2 TEMAS DSS semi-natural model

The semi-natural model implemented in the TEMAS DSS is the IAEA forest model, FORM (IAEA, 2002). It has been specifically developed to quantitatively estimate the contamination of temperate and boreal forests in the long-term after radionuclide deposition. In addition it has been designed to assist the selection of the best countermeasure based on its economic and dosimetric impact.

The model is divided into three modules (Figure 2.9): a radioecological model, which is used to estimate the radiocaesium activity concentration in semi-natural food products; a dosimetry model, implemented to assess the dose to people who spend time in the forest; and finally a countermeasure and economic impact model, which on the basis of the results from the two previous models, selects the best countermeasure considering the dose reduction and economic impact.

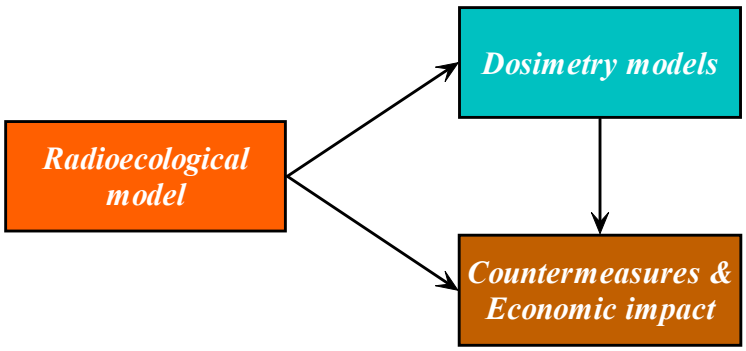


Figure 2.9 FORM structure. FORM includes three components: the radioecological model, the dosimetry assessment model and the countermeasure and economic impact model.

For the purpose of this work the radioecological model is the only FORM module which is assessed and therefore described in this section. For a detailed description the relevant literature is suggested (IAEA, 2002).

The radioecological model is formed by five dynamic compartments, which estimate the ^{137}Cs inventory in the main forest pools: leaves, bark, litter, organic matter and mineral soil layers (Figure 2.10)

The Leaves compartment (equation 2.3) accounts for the part of the tree surfaces that are directly contaminated by the initial radionuclide deposition. It loses its radionuclide load rather quickly, therefore it is an important compartment only in the short-term after deposition. Leaves and needles of deciduous and coniferous forests trees are included in this compartment.

$$\frac{dC_1}{dt} = I_2 - \left(R_{12} + \lambda\right) C_1 \qquad (2.3)$$

where

I_2 is the ^{137}Cs fraction deposited on the canopy

R_{12} is a first order rate coefficient (y^{-1}), (Table 2.2)

λ is the rate coefficient of radioactive decay

The bark compartment (equation 2.4) accounts for the direct contamination of the tree surfaces, i.e. bark. It is an important compartment in the short-term after the contamination, due to weathering and bark fallout the contaminant is rapidly passed to the soil surface layer. The ecological mean residence time in this compartment is set to 0.25 years for deciduous forest and 2 years for a coniferous forest.

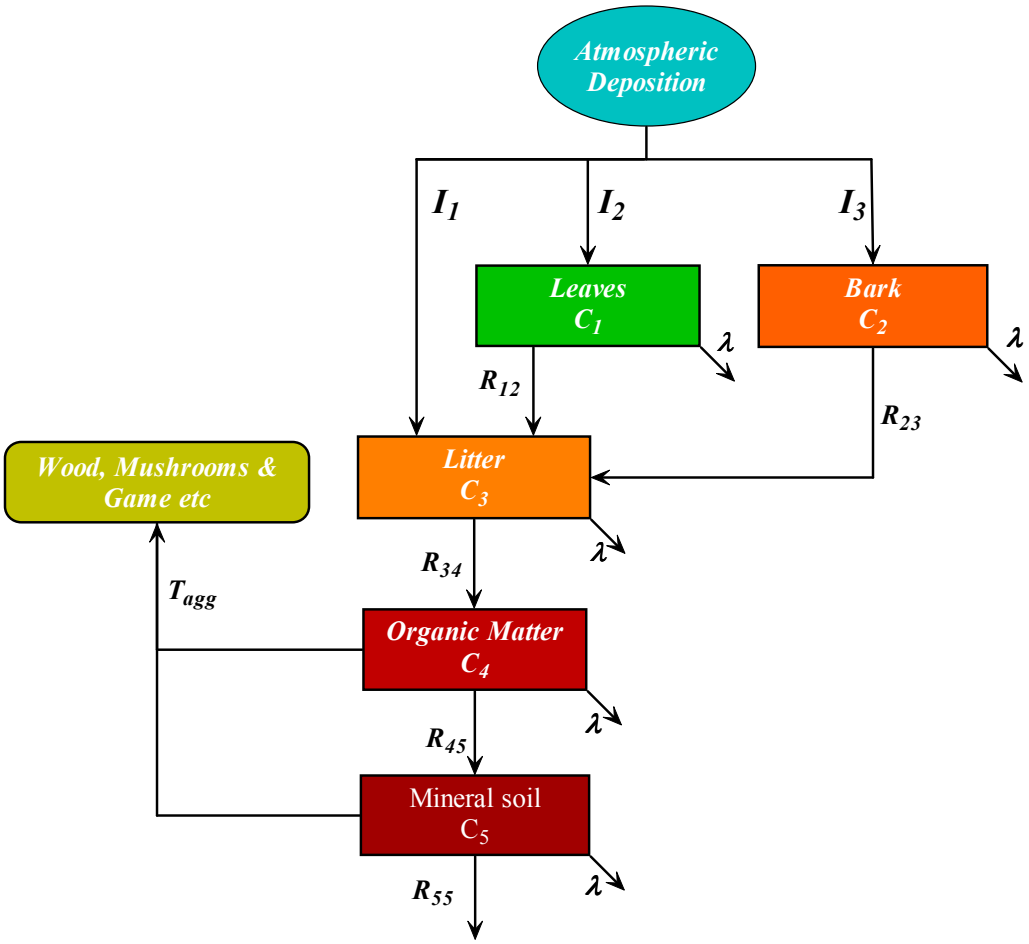


Figure 2.10 Structure of the FORM radioecological model.

$$\frac{dC_2}{dt} = I_3 - \left(R_{23} + \lambda\right) C_2 \qquad (2.4)$$

where

I_3 is the ^{137}Cs fraction deposited on the tree bark

R_{23} is a first order rate coefficient (y^{-1}), (Table 2.2)

λ is the rate coefficient of radioactive decay

The radionuclide activity concentration in semi-natural products is not estimated using a dynamic compartment as it is assumed to be in equilibrium with the mineral and organic soil layers.

Aggregated transfer factors (T_{agg}) are used to calculate the transfer of radionuclides from the mineral and organic soil compartments to a specified forest product (*SmPs*) (equation 2.5). These T_{agg} values relate activity inventories within organic and mineral soil layers, i.e. C4 and C5, (expressed in Bq/m^2) to that in the food product at equilibrium, (Bq/kg).

$$SmPs = T_{agg}(C_4 + C_5) \quad (2.5)$$

The litter compartment (equation 2.6) includes the part of the soil surface that is composed primarily by organic matter which is not yet decomposed. It includes leaves, branches and bark that fall on the soil due to natural causes or normal forest management.

The litter is directly contaminated from the initial deposition in the short-term, although in a smaller fraction than leaves and bark due to the canopy density which retains the large portion of the radionuclides. The main source of contamination received by the litter layer is due to the tree surface weathering and re-deposition of organic material falling from the canopy. Radionuclides in the litter layer are assumed not to be available for plant uptake.

$$\frac{dC_3}{dt} = I_1 + \left(R_{12} - R_{34} + R_{23} - \lambda\right)C_3 \quad (2.6)$$

where

I_1 is the ^{137}Cs fraction deposited on the litter soil layer

R_{12}, R_{34} and R_{23} are first order rate coefficients (y^{-1}), (Table 2.2)

λ is the rate coefficient of radioactive decay

The organic layer compartment (equation 2.7) represents the layer of organic matter decomposed from the litter layer. It is not directly exposed to the initial deposition and becomes contaminated through leaching from the litter layer. It is assumed to be one of the contributors to vegetation contamination due to the presence of plant roots.

The final compartment considered is the mineral compartment (equation 2.8), which simulates the contamination of the mineral soil layer, which is a source of contamination to plants due to root uptake.

$$\frac{dC_4}{dt} = \left(R_{34} - R_{45} - \lambda\right)C_3 \quad (2.7)$$

where

R_{34}, R_{45} are first order rate coefficients (y^{-1}), (Table2.2)

λ is the rate coefficient of radioactive decay

$$\frac{dC_4}{dt} = \left(R_{45} - R_{55} - \lambda\right) C_3 \tag{2.8}$$

where

R_{45}, R_{55} are first order rate coefficients (y^{-1}), (Table2.2)

λ is the rate coefficient of radioactive decay

Table 2.2 FORM parameters value

Parameter	Values (y^{-1})
R_{12}	0.086
R_{23}	0.35
R_{34}	0.46
R_{45}	0.046
R_{55}	0.01
λ	0.023

2.2.3 RIFE1

RIFE1 has been developed as part of the SEMINAT project (Belli et al, 1999) and aimed to increase the understanding of the behaviour of radiocaesium in semi-natural ecosystems.

RIFE1 is a dynamic model, which is formed by five compartments, representing the main radiocaesium pools in a forest ecosystem. The Cs fluxes between compartments, e.g. Cs percolation through the soil profile (Figure 2.11), are considered as first order rate processes and therefore represented by rate coefficient. Table 2.3 summarises the implemented coefficients.

RIFE1 can be divided into two parts: the above ground structure and soil horizons. Two compartments represent the above ground structure: tree external (equation 2.9) and tree internal (equation 2.10). Tree external simulates the tree bark contamination following the initial deposition; therefore it has a significant role in the tree inventory in the short term after the initial fallout. In contrast the tree internal compartment accounts for the Cs uptake by roots and foliar absorption which is

stored in the tree wood. In the long term this compartment represents the main biological pool of radiocaesium.

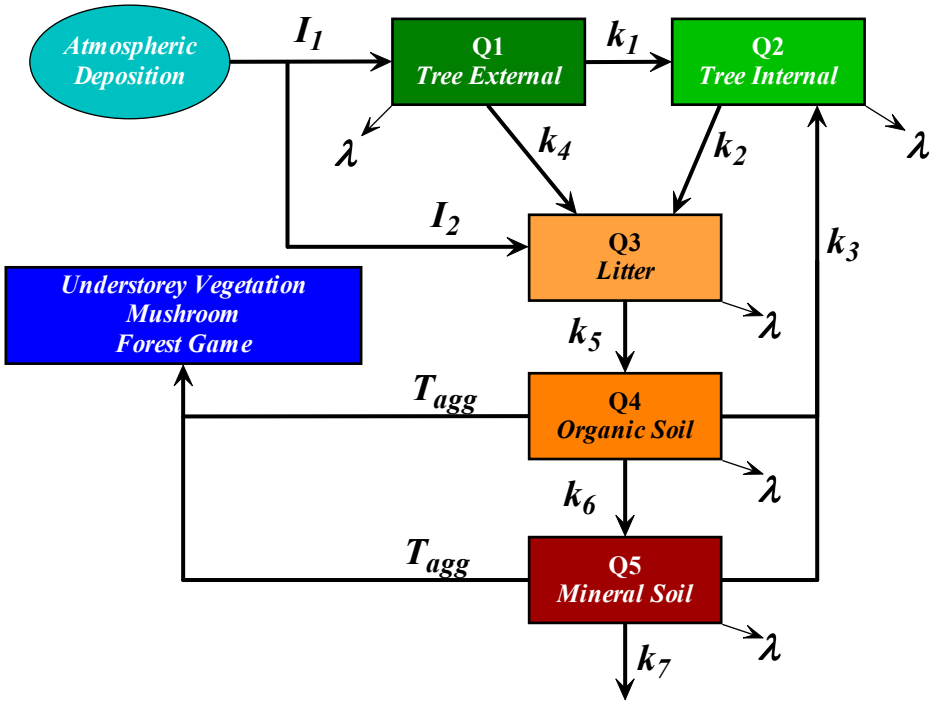


Figure 2.11 Structure of the RIFE1 model. Five compartments simulate the Cs inventory in the five ecosystem pools.

Table 2.3 Description of the rate coefficients used in the RIFE1 to simulate the Cs fluxes between ecosystem pools.

Symbol	Description	Units
k_1	Foliar absortion rate coefficient	y^{-1}
k_2	Foliar loss rate coefficient	y^{-1}
k_3	Tree root absortion rate coefficient	y^{-1}
k_4	Tree canopy weathering rate coefficient	y^{-1}
k_5	Rate coefficient of migration from soil litter layer	y^{-1}
k_6	Rate coefficient of migration from organic layer	y^{-1}
k_7	Rate coefficient of migration from mineral layer	y^{-1}
λ	Rate coefficient of radioactive decay	y^{-1}
T_{agg}	Soil-Understorey/Mushroom/Game transfer coefficient	m^2/kg

$$\frac{dQ_1}{dt} = I_1 - \left(k_1 + k_4 + \lambda\right)Q_1 \qquad (2.9)$$

$$\frac{dQ_2}{dt} = k_1 Q_1 + k_3 Q_4 - \left(k_2 + \lambda\right) Q_2 \qquad (2.10)$$

Where Q_1, Q_2 and Q_4 are the tree external, internal and the organic layer inventory (Bq/m^2) respectively. I_1 represents the fraction of Cs deposited on the tree external surface. k_n's are the rate coefficient (Table 2.3). Finally, λ is the rate coefficient of radioactive decay, i.e. 0.023 y^{-1} for ^{137}Cs or 0.33 y^{-1} for ^{134}Cs.

RIFE1 adopts a simple approach to describe the soil horizons, the soil profile is represented by three horizons: litter (Q_3), organic (Q_4) and mineral layers (Q_5), equation 2.11 to equation 2.13.

$$\frac{dQ_3}{dt} = I_2 + k_4 Q_1 + k_2 Q_2 - \left(k_5 + \lambda\right) Q_3 \qquad (2.11)$$

$$\frac{dQ_4}{dt} = k_5 Q_3 - \left(k_3 + k_6 + \lambda\right) Q_4 \qquad (2.12)$$

$$\frac{dQ_5}{dt} = k_6 Q_5 - \left(k_7 + \lambda\right) Q_5 \qquad (2.13)$$

Where Q_3, Q_4 and Q_5 are the Cs inventory for litter, organic and mineral layer respectively. I_2 is the fraction of Cs deposited on the soil surface. Finally, k_n's are the rate coefficient (Table 2.3).

The tree internal contamination is primarily due to the root uptake from the soil layers. An important feature of RIFE1 is the use of root distribution in the organic and mineral layers as weighting factors to estimate the Cs fluxes from the soil compartments and the tree internal.

The SEMINAT project collected measurements from seven European forests in the period between 1996 and 1998. The observations from these sites provided a "snapshot" of radiocaesium distributions in the forest compartments for a single year. RIFE1 has been calibrated for each site; therefore seven sets of model parameters are available, table 2.4.

Measurements of the k_2 value have been performed in most sites, however where such measurements were not available, i.e. Clogheen and Shanrahan sites, the

value obtained from Roundwood was adopted as it was considered to have similar tree characteristics (Belli et al, 1999). On the other hand k_4 values were obtained from the literature and every forest uses a value of 0.45 y^{-1}. In the case of k_3, it is almost impossible to obtain measurements for trees which are not part of a controlled experiment, therefore the value adopted is the result of a model fitting procedure. Estimation of k_1 is extremely difficult in a non-destructive way. Therefore the SEMINAT dataset does not include such observations and the literature does not provide reliable values. However, the process of Cs absorption from tree external surfaces to tree internal is assumed to have a negligible effect on the overall mass balance, therefore by default, k_1 is set to zero (Belli et al, 1999).

The Cs activity concentration in semi-natural products and foodstuff, e.g. mushrooms, is estimated using the T_{agg} values which are considered inputs parameters.

Table 2.4 Estimated values of RIFE1 half times obtained, by calibration, from the sites indicated

	k_2	k_3	K_4	k_5	k_6	k_7
Norvaggio	3.85	69.30	0.45	36.90	198.00	3.40E+03
Tarvisio	2.27	19.80	0.45	92.40	211.00	5.55E+02
Weinsberger	5.06	69.30	0.45	124.00	124.00	7.70E+03
Kobernausser	5.06	13.90	0.45	77.00	173.00	5.55E+03
Roundwood	2.31	6.93	0.45	18.50	220.00	3.47E+03
Clogheen	2.31	5.78	0.45	52.50	84.50	4.33E+03
Shanrahan	2.31	10.70	0.45	66.60	152.00	4.05E+03
Kruki	1.73	13.90	0.45	60.00	154.00	6.93E+02
Max	5.06	69.30	0.64	124.00	220.00	7.70E+03
Min	2.27	5.78	0.03	18.50	84.50	5.54E+01
GM	3.1	17.80	0.45	57.60	159.00	3.37E+03

2.3 SUMMARY

Following the Chernobyl accident several European countries were contaminated by radiocaesium (De Cort, 2001). Regarding terrestrial environments, i.e. rural and semi-natural, the main pathway, which affects human dose, is the ingestion of contaminated food products, e.g. milk and mushrooms.

The five models that have been considered in this work are summarised below.

SAVE DSS:

- SAVE rural model (Absalom et al, 2001): the soil-to-plant model uses a semi-mechanistic approach. The model inputs are: clay and organic matter content, exchangeable K^+ and pH. The output is the Cs activity concentration in plants. The Cs activity concentration in foodstuff is estimated using a transfer coefficient.
- SAVE semi-natural model: the model estimates the Cs activity concentration in semi-natural foodstuffs using a T_{agg}. Default values are present in the system, however the user can introduce values which resemble more accurately the scenario considered.

TEMAS DSS:

- TEMAS rural model: the radionuclide activity concentration in the edible and inedible part of the plant is estimated adopting an empirical model, which uses a database to access soil characteristics and TF's. Specific inputs are: exchangeable K^+ percentage of organic matter, clay and sand. Outputs: radionuclide activity concentration in edible and inedible part of the plant. The animal contamination is estimated using a transfer coefficient.
- TEMAS semi-natural model (FORM): The soil-to-plant interaction is simulated using a dynamic compartment model, on the other hand the semi-natural foodstuff contamination is estimated using a T_{agg} approach. The only input required is the T_{agg} value for the semi-natural product of interest.

RIFE1: The simulation of radiocaesium fluxes between ecosystem pools is performed using a dynamic compartment model. The estimation of the food product activity concentrations is performed using T_{agg} value that can be introduced by the user.

3. UNCERTAINTY AND SENSITIVITY ANALYSIS

Mathematical models are developed to approximate systems of different features and complexities. Many processes are so complex that physical experimentation is too time-consuming, expensive, or even impossible. As a result, mathematical models are often employed to explore such systems and processes; the reliability of the implemented models is essential.

Commonly a model is considered "reliable" if its predictions show a good agreement with empirical data; therefore the model performance is often assessed on the goodness of fit between predictions and measurements. While such tests are essential in model testing it should not be the only approach for the assessment of a model. In this chapter two aspects of model performance will be investigated: uncertainty on the model prediction and model sensitivity to variation in model factors, i.e. uncertainty and sensitivity analysis. This will be performed on the four radioecological sub-models introduced in chapter 2.

The estimation of the uncertainty on a model prediction is essential since it establishes the confidence interval around the predicted value, while the model sensitivity reflects the model structure, which can be compared to the system investigated. The analysis of model uncertainty and sensitivity helps to understand the differences between the conceptual and the constructed model.

The aim of this chapter is to estimate the prediction uncertainty of the five radioecological models and their sensitivity to parameters, i.e. inputs and model parameters.

The chapter comprises three sections: an introduction regarding the definition of uncertainty and sensitivity, a methodological section and finally a presentation and discussion of results.

3.1 INTRODUCTION

3.1.1 Model Parameter terminology

The literature does not use a standard terminology to refer to model inputs and parameters; this has often caused misinterpretations. In order to avoid such a confusion a brief overview of the terminology used in this work is provided.

There are two main categories of model parameters: site specific and generic, figure 3.1 (Monte et al, 1996). Site specific parameters are scenario specific and often correspond to model inputs, i.e. measurements, even described as local parameters. Nevertheless, models may require local parameters which cannot be properly measured, due to lack of knowledge or technological limitations, therefore they are estimated by model fitting and are referred to as optimised local parameters (OL parameters).

Generic parameters can be applied to any scenario and they can be divided into optimised global parameters (OG parameters) and model constants. OG parameters are obtained through some form of model fitting and are strongly influenced by the data and by the model design. On the other hand, parameters defined as constant are characterised by a value that is widely recognised by the scientific community as it has been proved to have a low level of uncertainty and variability, for instance radionuclide half-life.

The terms model factors and model parameters will be taken to mean the complete set of global and local parameters required by a model.

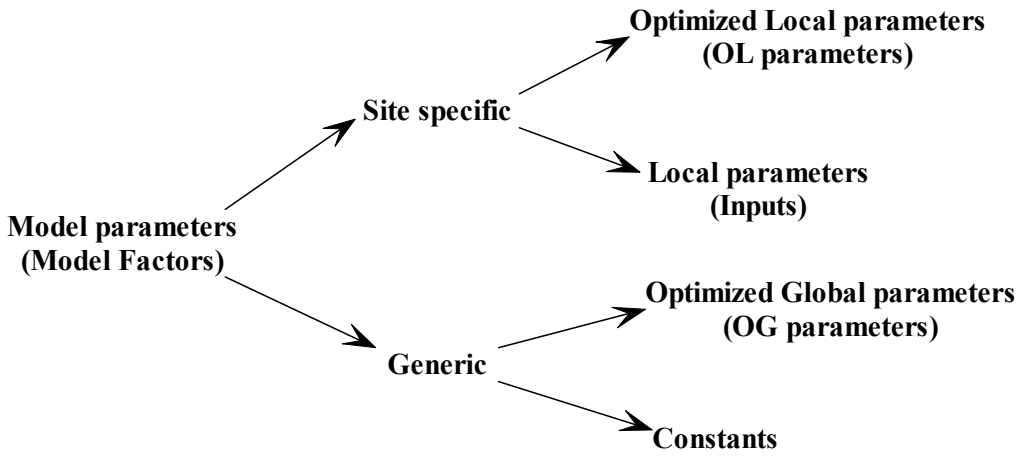


Figure 3.3 Model parameters are divided into site specific and generic parameters.

3.1.2 Uncertainty of model outputs

The Collins dictionary definition of uncertain is "*something that is not able to be accurately predicted*". Although this statement is quite general, it provides a good starting point for a discussion which is going to be developed more specifically in the following paragraphs.

The predictions of mathematical models are associated with some degree of uncertainty that affects model accuracy. The literature offers a wide range of uncertainty definitions and classification (e.g. Moschandeas et al, 2002; Oberkampf et al, 2002; Vardoulakis et al, 2002; Helton et al, 1995). Most of these are driven by the context to which they are applied, although there is a number of common features which can be used to classify the source of uncertainty into two categories: *parameters-scenario* and *uncertainty-variability*.

3.1.2.1 Parameters-scenario uncertainty

The *parameters-scenario* uncertainty classification suggests that the uncertainty on the model output is due to a combination of factors. Two categories have been defined: uncertainty associated with global parameters and uncertainty due to measurement error on the local parameters.

Parameter uncertainty refers to the uncertainty of optimised (i.e. fitted) parameters, which is generally due to the model design, commonly referred to as model structural error, and to the accuracy of the data set used for the model optimisation.

Scenario uncertainty refers to the uncertainty associated with the local parameters (empirical data), which describes the scenario investigated. The accuracy of such data is mainly due to measurement and aggregation error.

The measurement error is generally a function of two factors: the level of precision of the measurement technique and sampling strategy, which could produce random errors (imprecision), systematic errors (bias), or errors due to small sample size and non-representative samples (Moschandeas et al, 2002; Kuczera et al, 1998). Aggregation errors are due to spatial or temporal approximations following any homogeneity assumptions used to describe the scenario investigated.

3.1.2.2 Uncertainty-variability

The second approach to consider is the uncertainty-variability classification. Although the terms "variability" and "uncertainty" are often interchanged, a careful distinction should be made.

The variability, or *aleatory uncertainty* (Oberkampf et al, 2002), of a parameter or measurement represents the heterogeneity of the population considered. It cannot be reduced by a more accurate measurement or estimation, since it is part of the nature of the system. On the other hand, the uncertainty, or *epistemic uncertainty* (Oberkampf et al, 2002) is due to the poor or partial understanding of the driving processes or poor measurement performance (e.g. small sample size or inaccurate instrument). Potentially, it could be reduced.

3.1.2.3 Comparison between Parameters-Scenario uncertainty and Uncertainty-Variability

The literature considers the Parameters-Scenario uncertainty and Uncertainty-Variability as two distinct approaches to describe the uncertainty which affects models. However, they should be regarded as complementary and not as alternatives since they simply illustrate two different aspects of model uncertainty. Therefore the uncertainty of a model should be regarded as a combination of uncertainties on parameters, empirical data and model design and regarding empirical data and it should distinguish between uncertainty and variability. Therefore the concept of *aggregated model uncertainty* (AMU) could be introduced. However there is not a statistical test that could be used to estimate the AMU, consequently this concept might not find a practical application.

Therefore in the present work, the uncertainty on the model prediction has been considered to be due only to the combination of the uncertainties affecting model factors, which is the parameters-scenario approach.

3.1.2.4 Uncertainty analysis

The aim of the Uncertainty Analysis (UA) is to determine how uncertain is the output nominal value and how the uncertainty of model factors influences the model prediction, identifying and evaluating the probability density function (pdf) and the related statistics of the model outputs.

To perform such an analysis, it is essential that the model predictions are expressed in a probabilistic format, therefore a probability density function (pdf) needs to be established. The output uncertainty can be quantified by statistical parameters such as standard deviation (σ), and the use of the Coefficient of Variability or the Normalised Range at a specific probability (NR_p) could be adopted as an index to compare the uncertainty of different model predictions.

The methods that are suggested in the literature to evaluate the probability distribution of model outputs are the Monte Carlo Sampling (MCS) (Moschandeas et al, 2002; Papadopoulos et al, 2001; Crosetto et al, 2000; Helton et al, 1995) and Latin Hypercube Sampling (LHS) (Helton et al, 2005; Saltelli et al, 2000a; McKay, 1992).

MCS approach can be briefly summarised as following: random values are selected, for each variable considered, based on their probability distribution. The random selection process is repeated *n* times, and the model prediction is evaluated on each occasion. The N number of outputs corresponds to the prediction population, for which a probability density function (pdf) can be estimated and the model forecast confidence interval obtained. A limitation of this method is that the size of the sample *n* has a strong influence on the pdf produced. If the size of *n* is not sufficiently large, MCS might not evenly sample all the distribution space, in particular the distribution tails might not be considered.

The Latin Hypercube Sampling (LHS) has been developed to compensate the limitation of the MCS approach. Using the LHS, the pdf is divided into n_{LHS} intervals, having equal probability (figure 3.2), and *m* values are selected randomly from each of them, the x_{LHS} array represents the sample population (equation 3.1). Therefore, the LHS can be regarded as a more "efficient" method and it is preferred to the MCS when only a small sampling size *n* is computationally feasible, i.e. less than $1*10^3$ samples. Nevertheless, the models investigated in the present work are characterised by a fairly simple design, which implies that MCS can be successfully used, as a sufficiently large sampling size can be achieved i.e. $1*10^4$ samples.

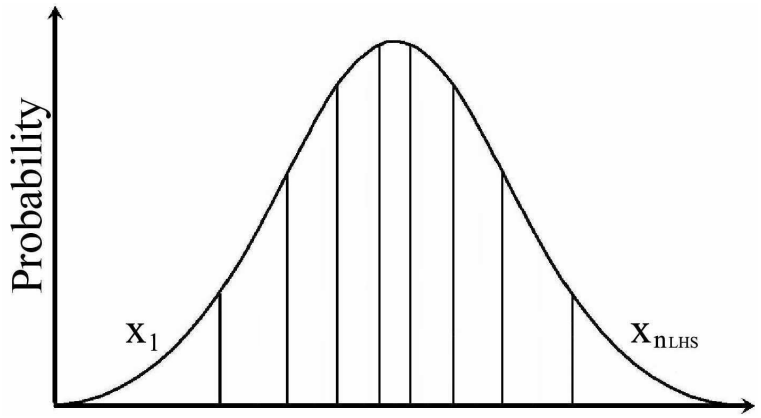


Figure 3.4 LHS method. The probability distribution is divided into regions having equal probabilities, values are randomly selected from each region.

$$x_{LHS} = \begin{bmatrix} x_{1,1} & \cdots & x_{n_{LHS},m} \end{bmatrix} \tag{3.1}$$

where:

n_{LHS} is the number of intervals used to divide the population pdf.

m is the number of samples selected from each interval.

3.1.3 Sensitivity

3.1.3.1 Introduction

Sensitivity analysis (SA) is the study of how variation in outputs can be qualitatively or quantitatively related to different components of the prediction uncertainty and how the given model output depends upon the local parameters. As a whole, SA is performed to increase the confidence in the model design and prediction since it establishes the fractional contribution of model factors to the uncertainty of the model predictions, therefore it can be considered as a complementary analysis to the UA. The outcome from such an analysis can be interpreted to answer three key questions:

(d) Does the model resemble the conceptual model?

(e) Which factors contribute the most to the output variation and may therefore require additional research?

(f) Which model parameters or variables can be considered insignificant, in the sense of not affecting the variation of the output and can therefore potentially be eliminated from the final model?

Under (a), the model may not appear to reflect correctly the dynamics involved in the process investigated, for example showing strong dependence to factors that should not be influential.

Under (b), SA could assist modellers in assessing the accuracy of the parameter estimation. Further work can be directed towards improving the estimation of those parameters, which would result in reducing the uncertainty of the model predictions.

Under (c), models should not be more complex than needed: factors or variables which do not affect the output variability could simply be removed. This process is generally referred to as the Parsimony Principle (Barnes, 2000).

Several methods have been developed and proposed to perform SA. The choice of which SA methods to adopt is a difficult task since each technique is characterised by strengths and weaknesses, therefore the whole decision process depends on the model characteristics and on the computational cost which can be afforded.

It is not the objective of this work to provide a complete overview of the SA proposed by the literature, consequently only a short summary of the most common approaches will be presented.

Saltelli et al (2000) propose a classification for SA methods: screening, local SA and global SA methods, each of these categories including several methods. For a comprehensive discussion on the SA methods presented in the literature, more detailed sources are suggested (e.g. Frey and Patil, 2002; Saltelli et al, 2000a; Turànyi 1990).

The difference between screening and local/global SA methods is that screening techniques are preliminary numerical experiments which can be used to identify the factors that control most of the output variability. The main advantage of these methods is that they require a low computational effort; however such "economical methods" have the drawback of providing only qualitative sensitivity measures. The considered factors are ranked in order of influence although a quantification of their weight is not provided. Local/global SA methods require larger computational effort and provide a quantitative estimation of the contribution of model factors on the variation of the model output.

In general terms in global SA techniques, parameters are varied simultaneously and the model sensitivity is measured over the entire range of the model factors variability; while in local SA methods, the parameters are changed one at a time and the factors variation is related to a small interval around the nominal value.

Local SA methods are suitable for investigating a limited region of space around the parameter nominal value. The variations on the nominal values are the same for every factor considered, which implies that the changes are not related to any knowledge of the model factors considered. In addition, local SA are mostly suitable for models where the parameters are characterised by comparable level of uncertainty.

On the other hand, global SA methods explore the entire space over which a factor may vary. Therefore the model sensitivity is estimated in respect of the factors uncertainty, which implies knowledge of the factors considered.

Global SA methods have the drawback that evaluation for complex models may be computationally prohibitive.

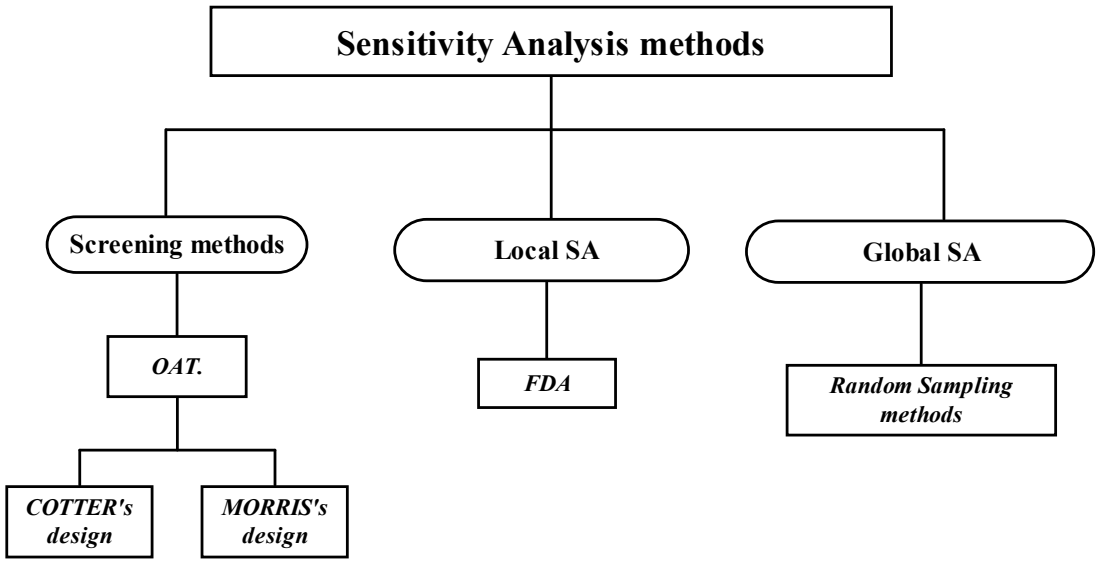


Figure 3.5 Three main types of Sensitivity analysis can be defined: local, global and screening.

3.1.3.2 Screening methods

The simplest case of screening methods is the one-at-a-time technique (OAT), where the impact of changing the values of each factor is evaluated in turn. Two extreme values are usually proposed to represent the range of a parameter likely value. The "midway" value of the range is considered to be the factor nominal value. The combination of the parameters nominal values is considered as the "control". The magnitude of the differences between the outputs for the "extreme" factors and the control are then compared in order to find the factors that significantly affect the model. Typically the number of evaluations required is $2k+1$, where k is the number of model factors.

A particular example of OAT is that proposed by Morris (Saltelli et al, 2000a). Unlike standard OAT, this has been considered as a global sensitivity experiment since it investigates the entire space over which the factors may vary. It is composed by individual randomised OAT tests, in which the impact of changing the value of each of the model factor is evaluated in turns.

The design developed by Cotter (Saltelli et al, 2000a) is a systematic fractional replicate design and is regarded as a local test. It requires $2k+2$ runs for a model characterised by k factors. The first run is performed with each factor at its lower level, afterwards k runs with each factor in turn at its upper level, while all the other k-1 factors have to remain at their lower level, following k runs with each factor in turn at its lower level, while all the other k-1 factors have to remain at their upper level and finally a run with every factor at their upper level. An array of output $y_0, y_1 \ldots y_k, y_{k+1} \ldots y_{2k}, y_{2k+1}$ is produced. Assuming that $S_o(j)$ and $Se(j)$ indicate the sum of the odd-order and even-order effects involving factor j respectively, the estimated values are $C_o(j)$ and $C_e(j)$, (equation 3.2 and 3.3). Finally the order of importance of each factor can be estimated using the measured $M(j)$, equation 3.4

$$C_o(j) = \tfrac{1}{4}\left[\left(y_{2k+1} - y_{k+j}\right) + \left(y_j - y_o\right)\right] \tag{3.2}$$

$$C_e(j) = \tfrac{1}{4}\left[\left(y_{2k+1} - y_{k+j}\right) + \left(y_j - y_o\right)\right] \tag{3.3}$$

$$M(j) = \left|C_e(j)\right| + \left|C_o(j)\right| \tag{3.4}$$

A drawback of the Cotter approach is that if a factor has effects which compensate each other, they may cancel each other out and the measures $C_o(j)$ and

$C_e(j)$ may not be successful. Although this event may be regarded as "unlikely", it is not impossible. Since there is no method to assess the occurrence of such a possibility, the results of this method should be critically assessed.

3.1.3.3 Local Sensitivity analysis methods

The simplest method to estimate the local model sensitivity is based on the OAT approach, changing one parameter at a time and re-evaluating the model. Through the application of the finite-difference approximation (FDA) the elements of a sensitivity matrix may be approximated by equation 3.5.

$$\frac{\partial y}{\partial k_j} \approx \frac{y\left(k_j + \Delta k_j\right) - y\left(k_j\right)}{\Delta k_j} \qquad j = 1 \ldots \text{m} \qquad (3.5)$$

where

y is the model output.

k_j is the m-vector of model factors.

This method is also called *brute force* method and it does not require any modification of the original model however it might be slow, since it requires m+1 simulations or 2m if central differences are considered and it is less accurate than more complex methods.

The accuracy of the estimated sensitivity is a function of the Δk_j adopted. In the case of nonlinear models, a Δk_j too large (e.g. >5%) would damage the initial assumption of nonlinearity. However, if the Δk_j is too small, the difference between the original and the perturbed y would not be appreciable. In most cases, a Δk_j equal to 1% is adequate, however determining a suitable value is a trial-and-error process for a given application (Saltelli et al, 2000a).

3.1.3.4 Global Sensitivity analysis methods

The main technique adopted to investigate the global sensitivity of a model is based on Sampling-Based methods, where the sampling process uses the random number approaches. Two main random sampling techniques can be identified: Monte Carlo Sampling (MCS) and Latin Hypercube Sampling (LHS), see 3.1.2.4 of this chapter. An important aspect of global sensitivity analysis is that they allow the

implementation of "correlation control" which enables us to consider correlations between model factors during the sampling.

Iman and Conover (Kendall et al, 1990) proposed a restricted-pairing technique to control the correlation of model factors in random sampling which is based on rank correlation rather than sample correlation. The implementation of this method ensures that the populations of random sampled values are characterised by the required correlation.

The fractional contribution of each parameter on the output variance may be determined by computing several regression analyses between the pdf of the model factors and the prediction pdf. The most common regression analyses techniques are: Standardized Regression Coefficients (SRC), Correlation Measures (Pearson), Partial Correlation Coefficients (PCC) and Rank Correlation Coefficient (RCC). Each of these methods is characterised by drawbacks and limitations in their application. As previously suggested, they are not going to be further discussed since it is not the objective of this work and because more relevant reading may be recommended (i.e. Saltelli et al, 2000a; Kendall et al 1990).

3.1.3.5 Conclusions

Global sensitivity analyses are reliable tests since they provide a quantitative estimation of the contribution, of each model factor, to the prediction uncertainty.

The MCS approach has the main limitation that it might be computationally expensive, however it is feasible for the present work as the models investigated can be considered relatively simple models.

3.2 METHOD

3.2.1 Software

The Monte Carlo sampling was undertaken using Crystal Ball 2000 Pro. (Crystal Ball, 1996), which is an Excel spreadsheet add-in tool designed for risk, uncertainty and sensitivity analysis.

The application of MCS requires the model factors to be described by a probability density function (pdf). In this project, global and local parameters have been described by a normal distribution. This is probably the most important

distribution in probability theory as it can be used to describe many phenomena and in addition, several statistical analyses can be applied only if the considered data are normally distributed.

To describe the uncertainty associated with model factors using a normal distribution (Figure 3.4), the population mean (μ) and standard deviation (σ) have to be known. For every model considered, except for RIFE1, the factor nominal value could be regarded as μ and the error on the nominal value was taken as σ. This approach could not be applied for RIFE1 as it is not implemented with any generic parameters. It has been independently calibrated for eight European forests; consequently eight different sets of optimised local parameters (OL) have been produced.

In the particular case of RIFE1 the distribution parameters (i.e. μ and σ) for the factors k_n have been evaluated considering the mean and standard deviation of the array k_n^m, where *n* is the number of OL parameters (1-7) and *m* is the number of data sets (1-8).

For the model factors where the error was not reported, 20% of the nominal value has been judged to represent a realistic estimation of a typical measurement error, therefore such a value has been adopted in the present work. Tables of the factors distribution values are reported in the section of this chapter where the uncertainty and sensitivity analysis results are presented for each model.

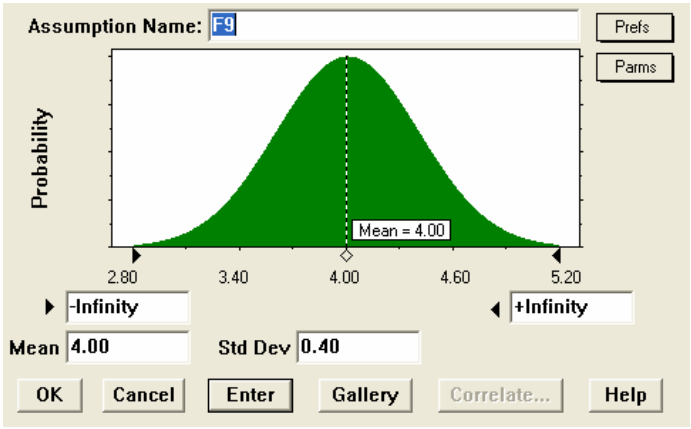


Figure 3.6 Crystal Ball Pro user interface for a Normal distribution setting.

The correlation between optimised model parameters has been estimated using the parameters covariance, which can be calculated during the model optimisation. In the present work the OpenModel software, developed at the University of Nottingham, has been used for this purpose.

3.2.2 Monte Carlo simulation method

The Monte Carlo method is described in Crystal Ball (1996) as follows: Considering a set of input data $a_1 \ldots a_m$, and a set of model parameters $p_1 \ldots p_m$, which are described by a probability distribution. The simulation process selects randomly one value for each variable based on its probability distribution. This process is repeated N times, resulting in N sets of model predictions, $Y_{(i)}$, i = 1 to N. The N outputs represent the prediction population. As a result, a probability distribution can be defined for the model prediction (**y**) (figure 3.5).

The correlation between model factors has been introduced using a specific Crystal Ball 2000 Pro. function, which implements the method proposed by Iman and Conover (Kendall, 1990) that requires the estimation of the Pearson Product-Moment Correlation Coefficient (*r*) between the considered factors.

The *r* can be defined as a measure of degree of linear relationship between two variables, equation 3.6, (Freund and Wilson, 2003) and it ranges between –1 and +1. The sign of the correlation coefficient defines the direction of the relationship, either positive or negative.

$$r_{xy} = \frac{\sum_{k=1}^{m}\left(x_k - \mu_x\right)\left(y_k - \mu_y\right)}{\left[\sum_{k=1}^{m}\left(x_k - \mu_x\right)^2\right]^{1/2}\left[\sum_{k=1}^{m}\left(y_k - \mu_y\right)^2\right]^{1/2}} \quad k = 1 \ldots m \qquad (3.6)$$

where:

m is the sample size.

x and *y* are the factors considered.

μ_x and μ_y are populations mean values.

The degree of correlation among model inputs can be determined using the PEARSON Excel function, which calculates the Pearson Product-Moment Correlation Coefficient (*r*) between two sets of measurements. On the opposite, the correlation between optimised parameters can be established during the optimisation process.

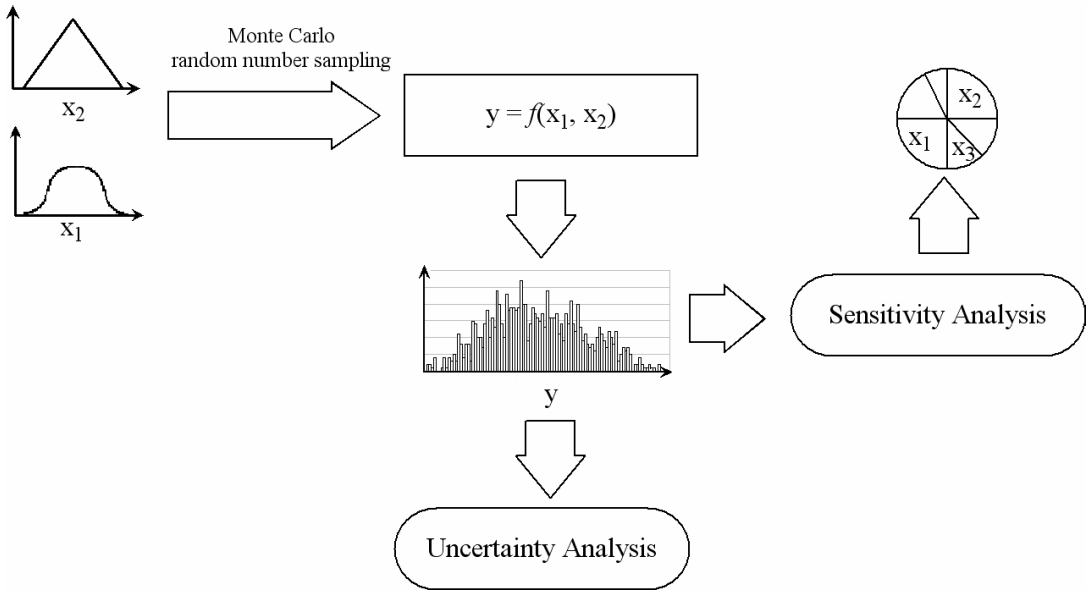


Figure 3.7 Uncertainty and Sensitivity analyses are performed considering the pdfs' estimated through the Monte Carlo sampling method.

3.2.3 Uncertainty comparison

The uncertainty on the model prediction can be quantified using the distribution standard deviation (σ), since it provides a measure of the distribution spread. However σ cannot be successfully adopted to compare the uncertainties of several model predictions, since it is not independent of units and scale. As a result the uncertainties associated with model outputs can be compared using the Coefficient of Variability (*CoeffVar*) (Freund and Wilson, 2003).

The *CoeffVar* is σ normalised by the distribution mean (μ), which makes it independent of units and scale, equation 3.7.

$$CoeffVar = \frac{\sigma}{\mu} \tag{3.7}$$

However, the coefficient of variability has the major drawback that it is a statistically representative measure of the distribution spread only when the variables considered are normally distributed. Consequently, for this project the Coefficient of Variability could not be adopted, as several model predictions are lognormal or exponentially distributed.

As a result of the limitation presented by the *CoeffVar,* a "pragmatic" index has been applied: the Normalised Range at a specific confidence interval of probability *p*, equation 3.8.

$$NR_p = \frac{Range_p}{\mu} \tag{3.8}$$

where:

μ is the population mean.

Range is the distribution range for a confidence interval of probability *p*.

From the statistical perspective, the distribution range could be regarded as a trivial parameter, however its use allows the index to be unaffected by distribution type, and it has been successfully applied in other works (e.g. McKay, 1992), however it can not be used to compare factors uncertainty as it is scale dependent, consequently a new index has been developed, equation 3.8. The NR_p is the distribution range of a confidence interval of probability *p* normalized by the population mean. As a result, this index can be successfully applied to compare the uncertainty on model factors.

In this project, a probability equal to 75% has been adopted to determine the NR_p. The selection of such a probability value has been based on the Absalom model uncertainty results, which will be discussed in more detail later in this chapter. The model output is exponentially distributed and the distribution range is extremely wide, 7 orders of magnitude at a probability of 95%, which represents the highest level of uncertainty among the models considered. The distribution tail is nevertheless characterised by a large number of values which have an extremely low probability. Therefore, a high number of trials would be required in order to assess their reliability. The use of 75% allows the analysis to focus only on that region of space where the values randomly generated can be considered accurate, without an excessive computational burden.

The method to determine the sampling size is a trial-and-error process. The accuracy of the pdf is positively correlated to the number of samples (*n*). However there is an optimum value for *n* (n_o) for which this correlation is valid, beyond this point increasing *n* does not affect the sampling accuracy on the considered pdf as it reached its maximum value (a_m), Figure 3.6.

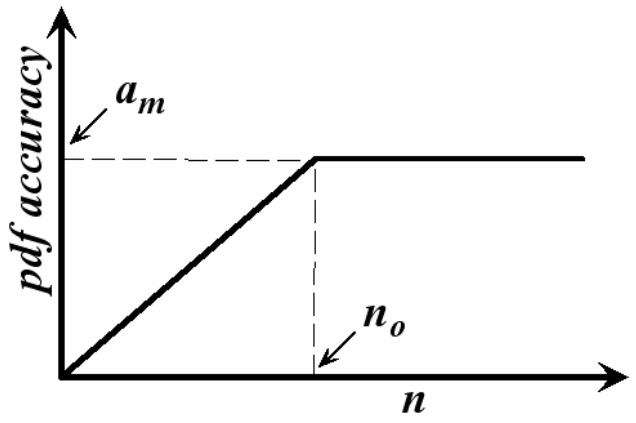


Figure 3.8 Correlation between pdf accuracy and the number of samples. The maximum level of accuracy (a_o) is associated to an optimum number of samples (n_o) .

To determine n_o the number of samples is progressively increased and the distribution X_{n+1} and X_n are compared using a Crystall Ball 2000 Pro function which estimates the difference between the two distributions. When the difference between X_{n+1} and X_n is close to 0, n is assumed to be equal to n_o

In the present work n was progressively increased in steps of one order of magnitude (i.e. 10^2, 10^3, 10^4 and 10^5). This test has been performed on one output of each model, and the optimum number of sample was determined to be 10^4 for each model.

3.2.4 Sensitivity analysis

Among the Regression analysis techniques, previously presented, the Rank Correlation Coefficient (RCC) has been adopted since it is the method available in Crystal Ball 2000 Pro (Crytal Ball, 1996). RCC is a useful approach which is often adopted in sampling-base sensitivity studies. It uses the standard Pearson correlation approach (r) (equation 3.6) to establish the relationship between inputs and outputs.

The Pearson correlation coefficient may not provide an accurate estimation if the relationship between inputs and outputs is not linear. In order to overcome this limitation, the rank transformation on the distribution is performed as part of the RCC. The rank transformation is a simple concept: the distribution data are first of all replaced by the appropriate "rank", and then, the Pearson correlation coefficient can be used to determine if a relation is present between inputs and outputs. The

ranks are assigned as following: the smallest value of each distribution is assigned number 1; the next largest value is assigned number 2 and so on up to the largest distribution number to which is assigned the number *m*, which is the total number of observations, (Saltelli et al, 2000a)

The use of the Rank transformation on the probability distributions has the drawback that information on the sample population is lost (Saltelli et al, 2000b). However, this is compensated by the fact that the RCC is *distribution-independent*, consequently it is possible to estimate the contribution of the inputs uncertainty on the output variance although the inputs are described by different distributions (Crystal Ball, 1996).

The presentation of the sensitivity analysis results through the use of RCC would not provide a clear answer to an important question: "what percentage of the prediction uncertainty is due to each specific model factor?". Although in this work the RCC has been used to estimate the sensitivity of the model predictions to model factor, the Contribution of Variance (*CV*) index (equation 3.9) had been preferred to the RCC to illustrate the sensitivity analysis results.

$$CV = \frac{100 * RCC^2}{\sum_{j=1}^{n} RCC^2} \qquad j = 1 \ldots n \qquad (3.9)$$

where

n is the number of factors considered.

The *CV* expresses the percentage of contribution for each model factor to the prediction uncertainty. However it is important to specify that this index can only be defined as an approximation of variance decomposition (Saltelli et al, 2000b).

3.3 RESULTS

3.3.1 SAVE - rural model sensitivity and uncertainty tests

The SAVE rural model (i.e. Absalom model) has been tested for uncertainty and sensitivity considering the effect of input and model parameters on the predicted soil to plant Transfer Factor (TF). Three "experiments" have been performed:

1. Determining the uncertainty and model sensitivity on three TF scenarios in order to assess changes in uncertainty and sensitivity with regards to different soil conditions.
2. Comparing the uncertainties and sensitivities at different time steps, 1, 10 and 20 years after initial contamination in order to estimate the evolution with time of such a model output.
3. Performing a "correlation test" to investigate the effects of considering the correlation between model factors on the UA and SA results.

In order to establish the effects of different soil conditions on the UA and SA results, the Absalom model has been implemented with measurements, which cover a wide range of soil conditions, from fifty-three different soils samples, derived from the Smolders (Smolders et al, 1997) and Sanchez (Sanchez et al, 1999) work (Table 3.1).

The highest (TF_{max}), middle (TF_{mid}) and lowest (TF_{min}) predicted TF have been selected and tested for uncertainty and sensitivity at three time points 1, 10 and 20 years after initial contamination. The model OG parameters, constant and local parameters used to estimate TF_{min}, TF_{mid} and TF_{max} are reported in table 3.2 and 3.3 respectively.

The TF_{min}, TF_{mid} and TF_{max} probability distribution function (pdf) for the three time periods are exponential (Figure 3.7a, b and c) and NR_{75} suggests that the model predictions are characterised by a rather high level of uncertainty (Table 3.4), however there are not major differences among them. This suggests that the uncertainty on the prediction is not caused by time or scenario characteristics, but is an intrinsic model feature.

Table 3.14 Soil characteristics of the Smolders (Smolders et al, 1997) and Sanchez (Sanchez et al, 1999) dataset.

	Smolders et al, 1997 (*Lolium perenne*)			*Sanchez et al, 1999* (*Agrostis capillaries*)		
	Min	**Median**	**Max**	**Min**	**Median**	**Max**
Clay (%)	0.50	12.25	31.30	2.00	7.60	57.60
OM (%)	3.50	6.15	34.00	12.60	75.20	96.50
pH	4.62	5.12	6.95	2.40	2.80	6.00
Exch.K (meq/100g)	0.15	0.62	1.84	0.10	0.40	0.80
RIP (mmol/kg)	54.00	1544.00	5861.00	5.00	157.00	6545.00
m_K (M)	7.1*10-5	9.9*10-4	1.2*10-2	3.0*10-5	1.1*10-4	1.7*10-3
m_{NH4}	0.00	4.5*10-5	6.4*10-3	0.00	7.8*10-4	4.7*10-3
TF	2.2*10-3	6.2*10-2	2.6	0.06	0.8*10-3	0.5*10-2

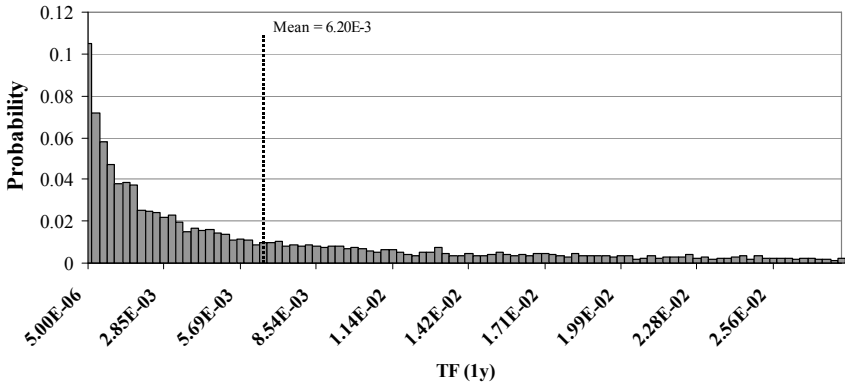
Table 3.15 Values of fitted and constants (*) used in the original model (Absalom et al, 2001). Each value is reported with the standard error (±).

Model parameters	Value
k_1	2.57 ± 0.1
k_2	1.56 ± 0.32
k_3^*	3.37
k_4^*	0.16
k_5^*	-34.66
k_6^*	29.72
k_7	4.18 ± 0.27
k_8	0.043 ± 0.017
k_9	1.74 ± 0.17
k_G^{clay}	3.18 ± 0.44
k_G^{humus}	2.32 ± 0.1
k_{fast}^* (1/d)	$1.9*10^{-3}$
k_{slow}^* (1/d)	$1.9*10^{-4}$
P_{fast}^*	0.814
CEC_{clay}^* ($cmol_c$/kg)	50

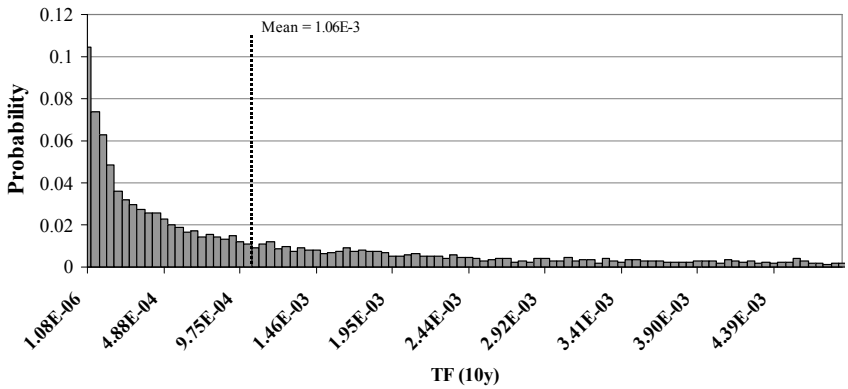
Table 3.16 The mean (μ) and standard deviation (σ) of soil characteristics for TF_{min} TF_{mid} TF_{max}.

	TF_{min}		TF_{mid}		TF_{max}	
	μ	**σ**	**μ**	**σ**	**μ**	**σ**

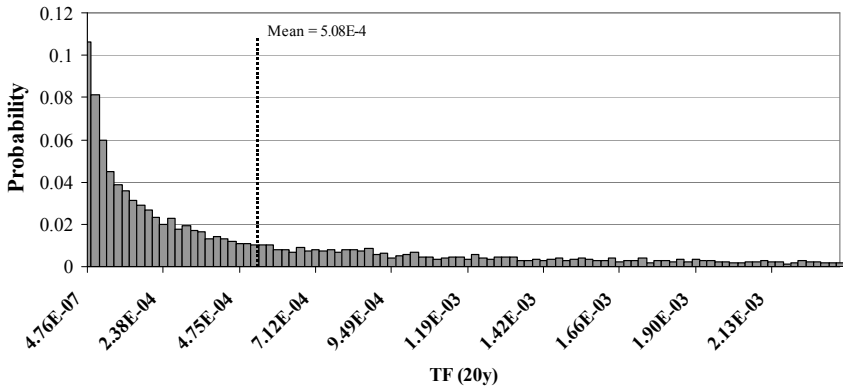
OM (g/g)	$9.50*10^{-2}$	$1.90*10^{-2}$	$2.61*10^{-1}$	$5.22*10^{-2}$	$9.52*10^{-1}$	$1.90*10^{-1}$
Clay (g/g)	$2.61*10^{-1}$	$5.22*10^{-2}$	$3.37*10^{-1}$	$6.74*10^{-2}$	$4.50*10^{-2}$	$9.00*10^{-3}$
PH	6.95	1.39	3.00	$5.80*10^{-1}$	3.01	$5.20*10^{-1}$
K^+ (cmol/kg)	1.27	$2.53*10^{-1}$	$4.00*10^{-1}$	$8.00*10^{-2}$	$4.00*10^{-1}$	$8.00*10^{-2}$
m_{NH4} (mol/l)	$1.06*10^{-4}$	$2.12*10^{-5}$	$7.80*10^{-4}$	$1.56*10^{-4}$	$3.50*10^{-4}$	$7.00*10^{-5}$



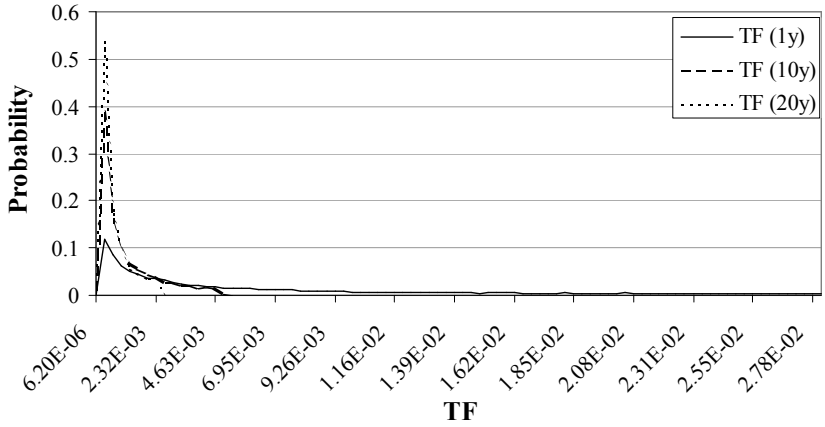
(a)



(b)



(c)



(d)

Figure 3.9 Probability density functions of TF_{min} for 1, 10 and 20 years after contamination, a, b and c respectively. The distribution represents the 75% confidence interval. The three distributions are compared in the figure 3.7d. The distribution of TF_{max} , TF_{med} is not reported as it is similar to TF_{min}.

Table 3.17 NR_{75} estimated for the three TF categories at 1, 10 and 20 years after contamination.

	NR_{75}		
	1 (y)	10 (y)	20 (y)
TF_{max}	6.38	6.84	6.89
TF_{med}	6.12	6.01	6.12
TF_{min}	4.59	4.58	4.67

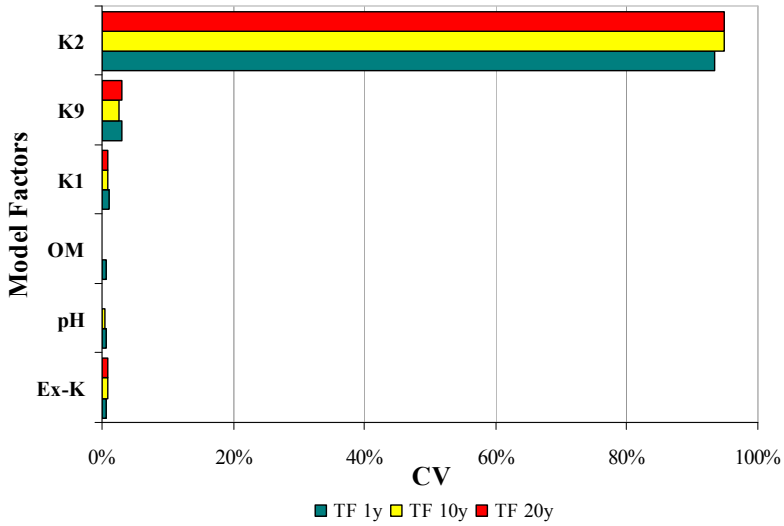


Figure 3.10 TF_{min} sensitivity analysis results. Results of the SA for TF_{max} and TF_{med} are not reported as they are similar to TF_{min}.

Table 3.18 Estimated NR_{75} for TF_{min}, TF_{med} and TF_{max} considering the model factors correlation.

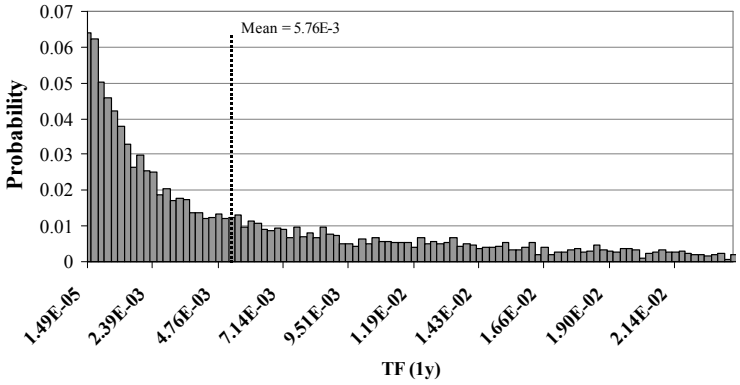
	NR_{75}		
	1 (y)	**10 (y)**	**20 (y)**
TF_{max}	6.04	6.40	6.18
TF_{med}	5.52	5.34	5.46
TF_{min}	4.12	4.12	4.15

The results of the SA (Figure 3.8) illustrate that the model sensitivity is entirely dominated by the k_2 parameters, which is a global parameter used in the CF sub-model. The contribution of k_2 is greater than 90% on the predicted TF uncertainty. The other model factors, including the local parameters, have a secondary role. The high model sensitivity to a single parameter could be considered as the cause of the rather high uncertainty observed on the predicted TF.

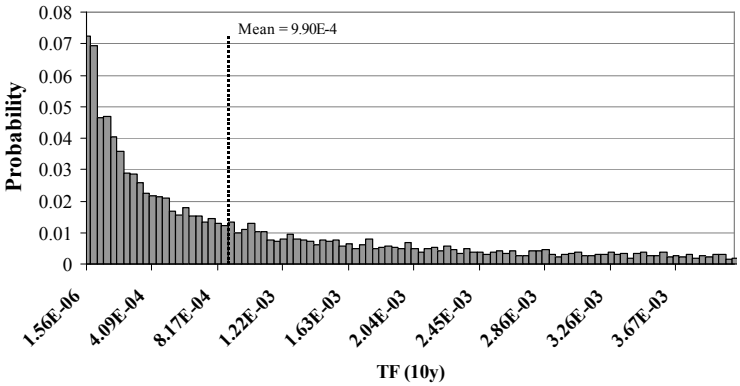
The final experiment performed on the Absalom model aims to establish the importance of the correlation between model factors during the MCS.

In order to determine the degree of correlation between local parameters, i.e. model inputs, a correlation matrix has been obtained for the five soil characteristics, i.e. OM, clay content, pH, K^+, and m_{NH4}. The results suggest that the only soil characteristics that have a significant level of correlation are the organic matter and pH, *r* equal to –0.70.

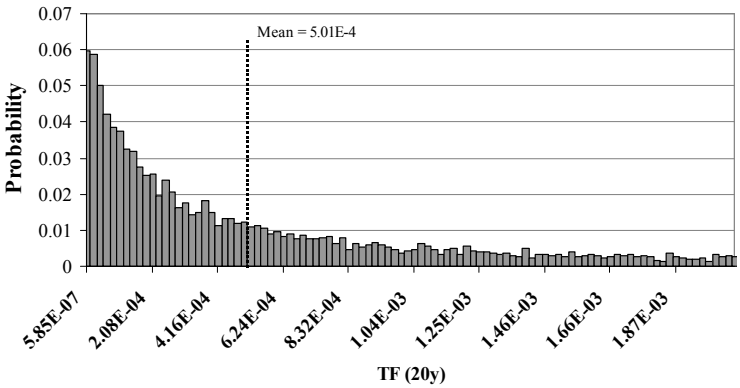
On the other hand the correlation among global parameters have been estimated using the OpenModel software. A significant correlation has been identified for k_2 and k_1, *r* equal to 0.96.



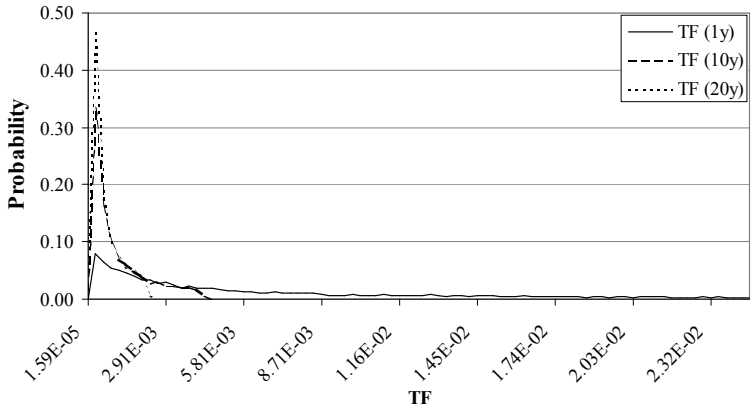
(a)



(b)



(c)



(d)

Figure 3.11 TF_{min} probability density functions at 1, 10 and 20 years, a, b and c respectively, considering the correlation among model factors. The Three pdfs' are compared in figure 3.9d. Due to the similarity of the results, the TF_{max} and TF_{med} are not reported.

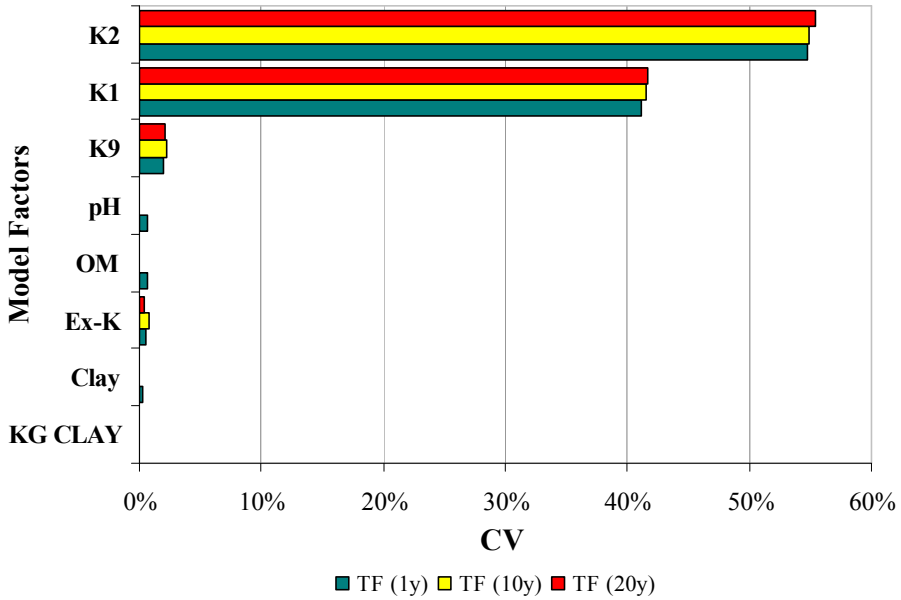


Figure 3.12 TF_{min} sensitivity analysis results taking into account the model factors correlation.

Only a slight decrease in the uncertainty on the predicted TF is detectable, table 3.4 and 3.5, if the correlation among model factors is included in the MCS sampling procedure.

On the other hand the sensitivity analysis suggests that the output uncertainty is equally influenced by the k_1 and k_2 parameters, 45% and 50% respectively, which are the two global parameters used by the CF sub-model. This result was expected since the *r* value considered is very high.

Although this result is different from that presented earlier, it leads to an equivalent conclusion that highlights the key role of the CF sub-model into the Absalom model.

3.3.1.1 Conclusions

The Absalom model is considered a semi-mechanistic model since it estimates plant contamination, simulating several chemical processes that take place in soil. Therefore the main objective of this model was to have a flexible system that could be applied successfully on a wide range of scenarios. However there is a substantial difference between the conceptual and the constructed model, since the model does not appear to be sensitive to local parameters as it was expected to be. In fact the uncertainty is entirely dominated by the CF sub-model (i.e. k_2 and k_1). This might be a system limitation if the model optimisation is not performed considering a wide range of soil conditions.

The issue of the low influence of several local parameters (i.e. pH) on the model prediction uncertainty could represent one of the major limitations of this model; therefore two different processes should be undertaken:

- Model simplification. The model is simplified in order to consider only the local parameters which have an effect on the model prediction. The reduction of model complexity might decrease the uncertainty on the prediction.
- Model re-elaboration. An evaluation of the model, in particular the CF sub-model, should allow to reduce the influence of k_2 parameter and therefore to establish a more balanced influence between local and global parameters. Such an approach is going to be further discussed in the fourth chapter of this thesis.

3.3.2 TEMAS sensitivity and uncertainty tests

The TEMAS rural model is characterised by a unique design among the models considered. It uses data, such as organic matter, sand and clay content, as input to classify the scenario under investigation into four soil categories: organic, mineral sand, mineral loam and mineral clay. Each of the soil categories is associated

with a data set of scenario characteristics (for instance average annual rainfall, soil porosity and k_d).

This model design reduces the number of local parameters a user has to provide and thereby increases the practical applicability of the model. However it is essential that the soil categorisation describes homogeneously the entire range of soil characteristics which could be encountered.

As a result, the robustness of the model design has been tested in addition to the uncertainty and sensitivity of the predicted TF. Two different strategies have been applied:

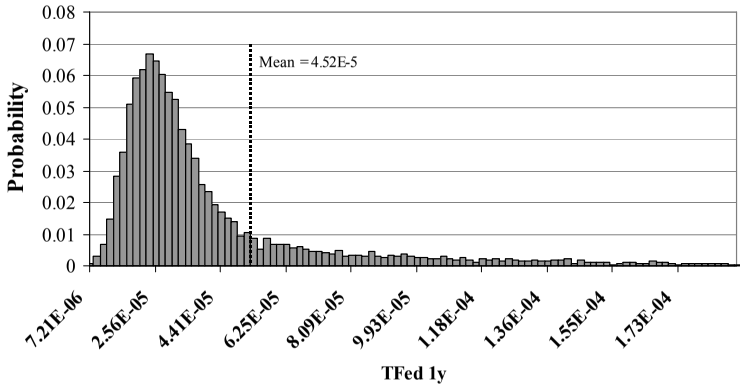
1. Performing the tests on a single soil scenario in order to estimate the uncertainty and the sensitivity of the model prediction (i.e. similar to that described for the Absalom model).
2. Considering the four soil categories to test the reliability of such soil categorisation.

Both tests have been performed on the edible (TF_{ed}) and inedible part of the plant (TF_{ined}) at two time points, 1 and 10 years after contamination.

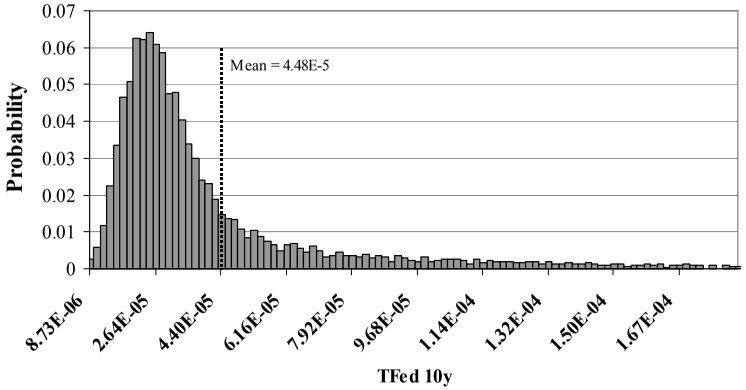
The correlation between model factors has not been considered, as an appropriate dataset, which would have been essential to estimate *r*, was not available. However, this has not been regarded as a limitation in the analysis since the results of the "correlation test", performed on the Absalom model, suggest that not considering model factor correlation does not have a significant impact on the outcome of the UA and SA.

3.3.2.1 Single soil category analysis

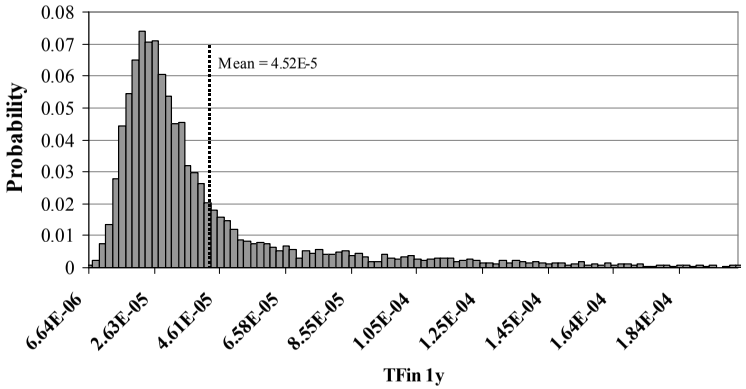
The pdf evaluated for TF_{ed} and TF_{ined} (Table 3.6 and 3.7) at the two time points are lognormal (Figure 3.11) and they are characterised by a low level of uncertainty (Table 3.8), NR_{75}>1.5. There is no significant variation between the prediction uncertainty at 1y and 10y.


Probability
Mean = 4.52E-5
7.21E-06
2.56E-05
4.41E-05
6.25E-05
8.09E-05
9.93E-05
1.18E-04
1.36E-04
1.55E-04
1.73E-04
TFed 1y

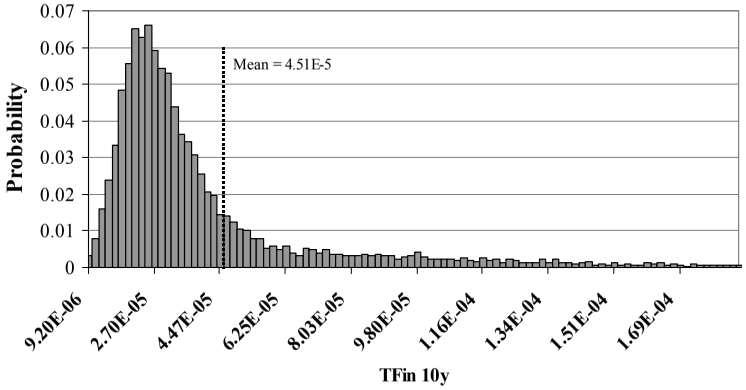
(a)


Probability
Mean = 4.48E-5
8.73E-06
2.64E-05
4.40E-05
6.16E-05
7.92E-05
9.68E-05
1.14E-04
1.32E-04
1.50E-04
1.67E-04
TFed 10y

(b)


Probability
Mean = 4.52E-5
6.64E-06
2.63E-05
4.61E-05
6.58E-05
8.55E-05
1.05E-04
1.25E-04
1.45E-04
1.64E-04
1.84E-04
TFin 1y

(c)


Probability
Mean = 4.51E-5
9.20E-06
2.70E-05
4.47E-05
6.25E-05
8.03E-05
9.80E-05
1.16E-04
1.34E-04
1.51E-04
1.69E-04
TFin 10y

(d)

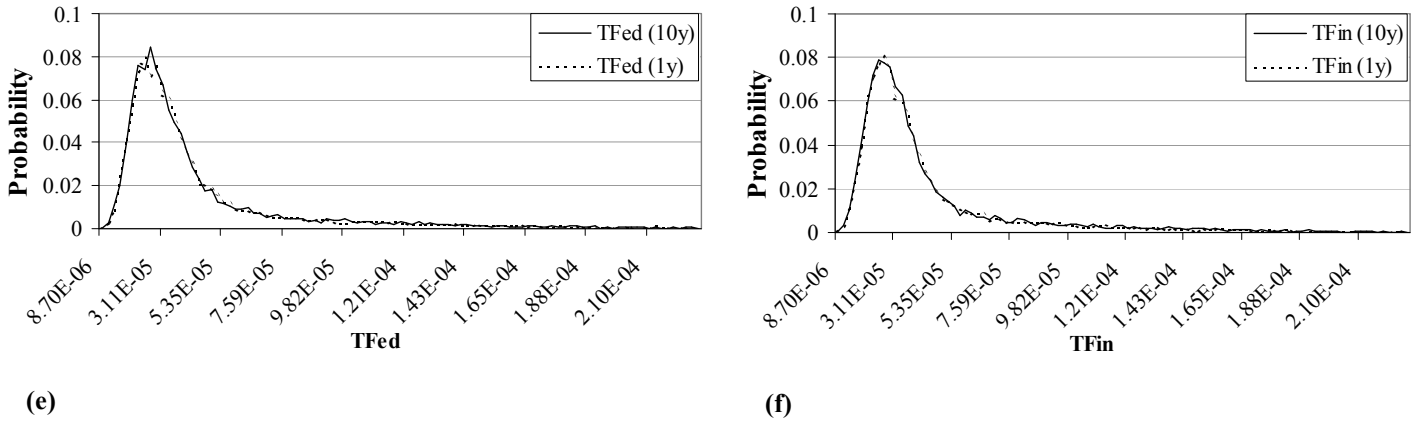


Figure 3.13 Probability density function, at 99 % confidence interval, estimated for TF_{ed} (a,b) and TF_{ined} (c,d) at 1 and 10 years after deposition. The two sets of pdfs' are also compared (e,f).

Table 3.19 Soil characteristics. The mean (μ) and standard deviation (σ) are reported.

	μ	σ
OM (%)	5	1
Clay (%)	15	3
Sand (%)	60	12
K^+ (cmol/kg)	0.6	0.12

Table 3.20 TEMAS database parameters. The mean (μ) and standard deviation (σ) are reported.

	μ	σ
Annual rain (mm)	886	177.2
Fibre Density	900	180
Soil mineral density	2650	530
Org. density	1400	280
Porosity (%)	0.48	0.096
k_d (dm^3/kg)	4400	880
Upper-TF (g/g)	1.2	0.24
Recom-TF (g/g)	0.12	0.024
Kexch_min (g/g)	0.5	0.093

Table 3.21 NR_{75} for TF_{ed} and TF_{ined} at the two considered time periods.

	NR_{75}	
	1y	**10y**
TF_{ed}	1.23	1.24
TF_{in}	1.22	1.20

The comparison between the SA results, performed at the two time points, shows that the system sensitivity does not have any considerable variation with time (Figure 3.12a and b).

Although the low uncertainty of the prediction is desirable, the sensitivity analysis performed on the two model outputs highlights a potential weakness in the system. The predicted TF in the edible and inedible part of the plant is mostly sensitive to soil and scenario characteristics that are not formally input parameters because they are part of the TEMAS database, (Figure 3.12a and b). This represents a major limitation in the model design, as the accuracy of the TEMAS predictions is

entirely based on the quality of the database, therefore this model may provide an adequate prediction only if the data included in the database provide a representative estimation of the characteristics of the scenario considered.

The database has to be carefully assessed, since it represents the key element of the model and it can strongly affect the final model prediction and its implementation.

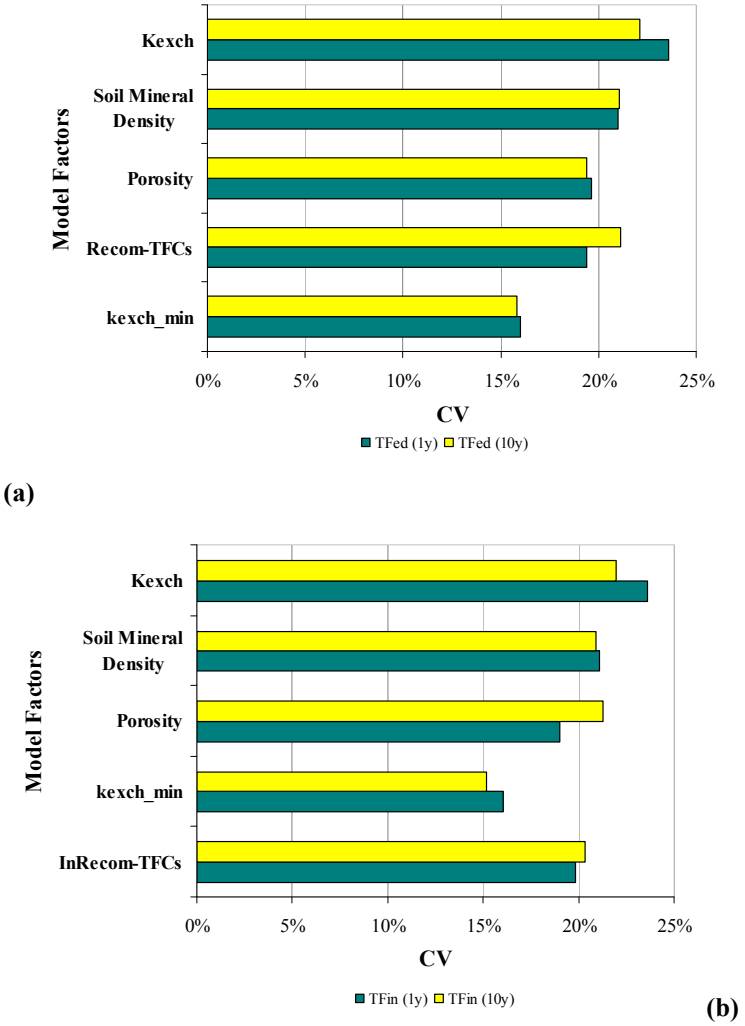


Figure 3.14 Sensitivity analysis results for TF_{ed} (a) and TF_{ined} (b) at 1 and 10 years.

3.3.2.2 Test on the four soil categories

The soil categorisation approach adopted in TEMAS has the advantage of reducing the number of input parameters which the user must supply, since the database provides the model with a large set of data. The drawback of such a method is that the four soil categories may not adequately describe every possible scenario. The following analysis aims to estimate the robustness of this approach.

As previously, model factors are described using a normal distribution. However, a discrete uniform distribution has been used to describe the database parameters for two reasons. In a uniform distribution, every value has the same probability to be selected, Figure 3.13. In addition to that, a discrete uniform distribution allows only the selection of the values which describe the four scenarios included in the TEMAS database, Table 3.9.

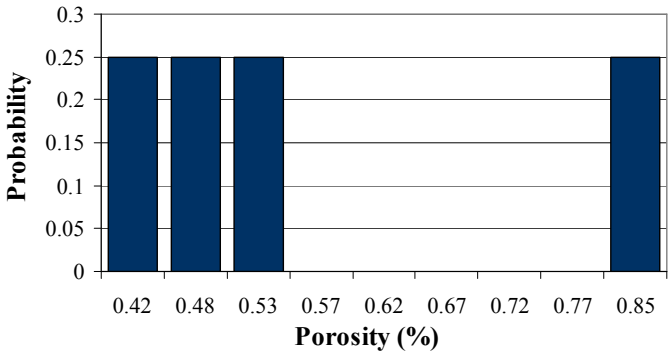


Figure 3.15 Discrete Uniform Distribution of the soil porosity for the four scenarios described in the TEMAS database.

Table 3.22 The TEMAS database parameters values for the four scenarios considered are reported below.

	Database parameters value			
	Organic	Mineral sand	Mineral loam	Mineral clay
Mean annual Rain (mm)	$1.22*10^{+3}$	$8.86*10^{+2}$	$1.17*10^{+3}$	$1.01*10^{+3}$
Porosity (%)	$8.50*10^{-1}$	$4.20*10^{-1}$	$4.80*10^{-1}$	$5.30*10^{-1}$
Kd (dm^3/kg)	$2.70*10^{+2}$	$2.70*10^{+2}$	$4.40*10^{+3}$	$1.80*10^{+3}$
Recom-TF (g/g)	$2.90*10^{-1}$	$2.10*10^{-1}$	$1.20*10^{-1}$	$6.60*10^{-2}$
Upper-TF (g/g)	5.50	1.70	1.20	$5.80*10^{-1}$
InRecom-TF (g/g)	$2.90*10^{-1}$	$2.10*10^{-1}$	$1.20*10^{-1}$	$6.60*10^{-2}$
InUpper-TF (g/g)	5.50	1.70	1.20	$5.80*10^{-1}$

The expected outcome of this test is that the four scenarios should be described by an equivalent number of distributions which show a degree of overlapping, figure 3.14. In addition the distribution means should have a comparable probability. However, the pdfs' evaluated for TF_{ed} and TF_{ined} demonstrate that the database approach does not cover homogeneously the entire range of possible scenarios. As the figure 3.15 a-d shows, some scenarios have a higher probability of occurrence than others and the degree of overlapping between distribution is low or null, creating "gaps" which represent scenarios with a probability of occurrence extremely limited.

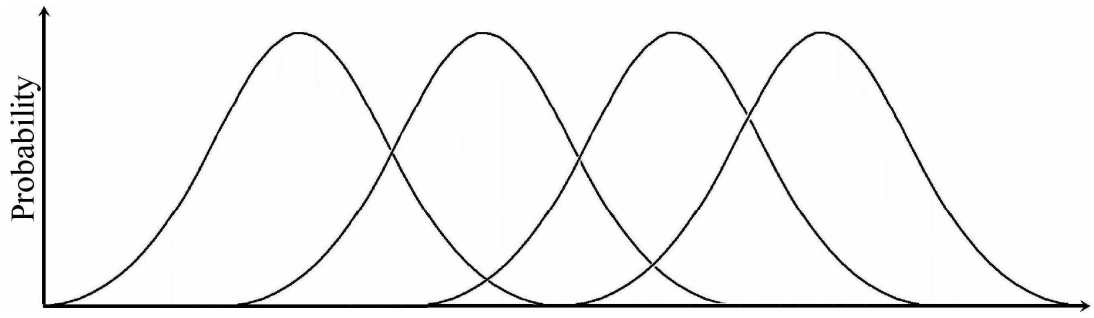


Figure 3.16 Expected outcome of the TEMAS database test.

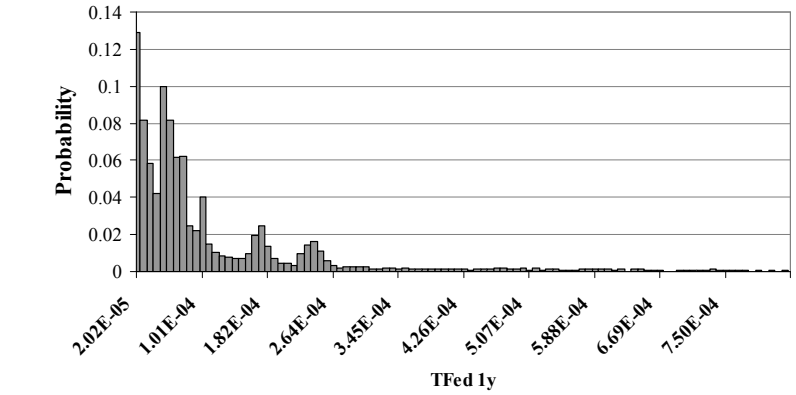
3.3.2.3 Conclusions

The predicted TF for the edible and inedible part of the plant are characterised by a desirable level of low uncertainty, however the model is mostly sensitive to factors included in the TEMAS database, therefore the key issue is the reliability of these data.

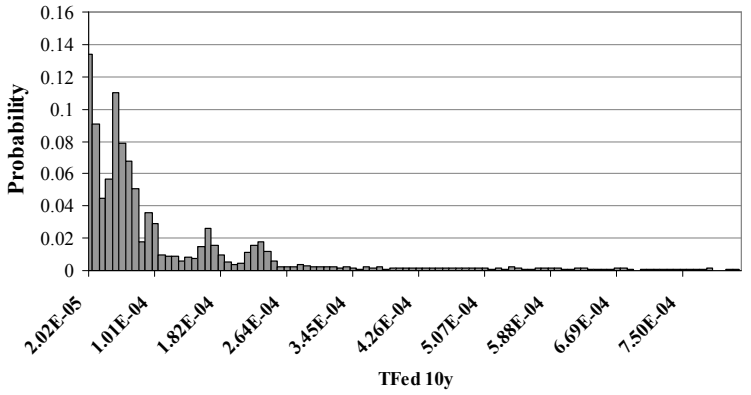
In addition, despite the fact that the database approach reduces the number of inputs supplied by the user, the test on its robustness suggests that the database does not offer a high level of flexibility, since it is not able to represent a wide range of scenarios, which is illustrated by the inhomogeneous TF_{ed} and TF_{ined} pdf. This model weakness might strongly influence the applicability of TEMAS in a decision making exercise.

Regarding the differences between the conceptual and the constructed model, It is suggested that the limitation in the model design, represented by the database should not be considered a discrepancy between the two. The use of a database,

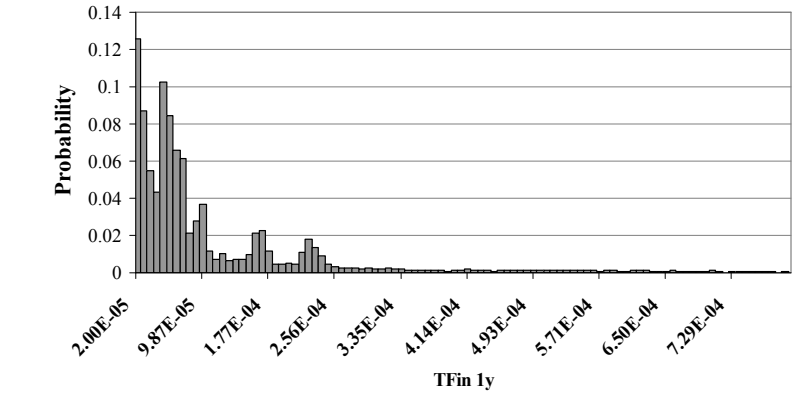
which incorporates several key scenario characteristics and the related soil classification, produces two obvious side effects: a system which is not particularly flexible and that some potential scenarios are not adequately considered. One can conclude that the constructed model is in agreement with the conceptual model, however, it is the conceptual model which has been based on weak assumptions.



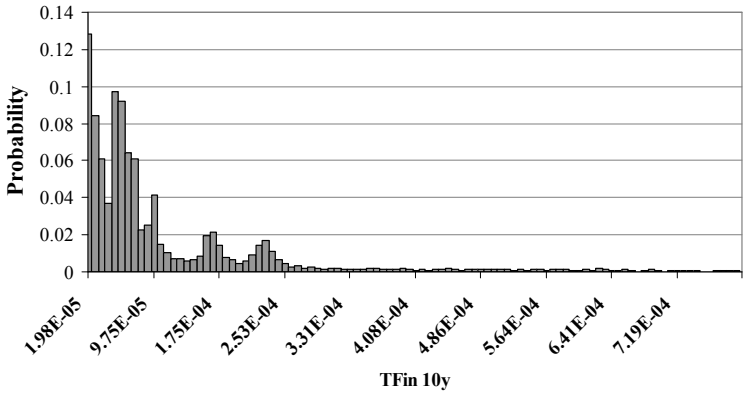
(a)



(b)

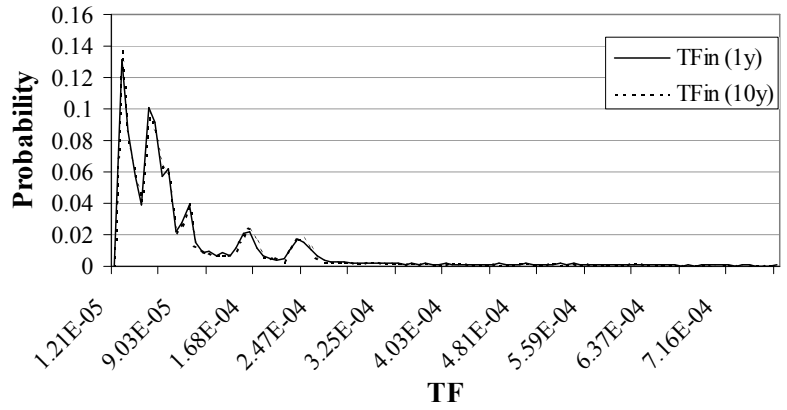


(c)

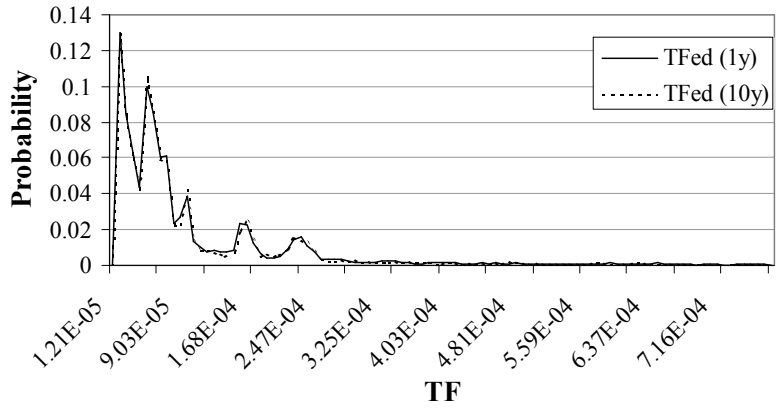


(d)

Figure 3.15 Soil probability density functions at 1 and 10 years after contamination. Discrete uniform samples result in multimodal distribution of TF.



(e)



(f)

Figure 3.15 Soil probability density functions at 1 and 10 years after contamination. Discrete uniform samples result in multimodal distribution of TF. The two sets of distribution are also compared, (e and f).

3.3.3 RIFE1 uncertainty and sensitivity analysis

RIFE1 has been tested for uncertainty and sensitivity on three of its outputs: ^{137}Cs activity concentration in bark (Bq/kg), mushroom (Bq/kg) and roe deer (Bq/kg). Two experiments have been performed:

1. Determining the uncertainty and model sensitivity on the three end-products, considering the Tarvisio dataset, table 3.10, for the global parameters and μ_n^m and σ_n^m, Table 3.11, to describe the pdf of local parameters.
2. Comparing the uncertainties and sensitivities at different time steps, 1, 10 and 20 years after initial contamination, to establish if time has an influence on the output uncertainty, Table 3.12.

The correlation between model factors has not been considered, during the MCS sampling, as measurements, which could be used to determine *r*, were not available.

Table 3.23 RIFE1 optimised local parameters for the eight European forests

	k_2	k_3	k_4	k_5	k_6	k_7
Norvaggio	3.85	69.3	0.45	36.9	198	3400
Tarvisio	2.27	19.8	0.45	92.4	211	555
Weinsberger	5.06	69.3	0.45	124	124	7700
Kobernausser	5.06	13.9	0.45	77	173	5550
Roundwood	2.31	6.93	0.45	18.5	220	3470
Clogheen	2.31	5.78	0.45	52.5	84.5	4330
Shanrahan	2.31	10.7	0.45	66.6	152	4050
Kruki	1.73	13.9	0.45	60	154	693
μ	3.13	20.09	0.45	66.43	151.25	4298.83
σ	1.35	26.96	0.00	32.74	45.89	2360.85

Table 3.24 Estimated NR_{75} for the RIFE1 outputs.

	NR_{75}		
	1y	**10y**	**20y**
Mushroom	0.26	0.21	0.16
Bark	1.33	1.25	1.31
Roe Deer	0.22	0.13	0.14

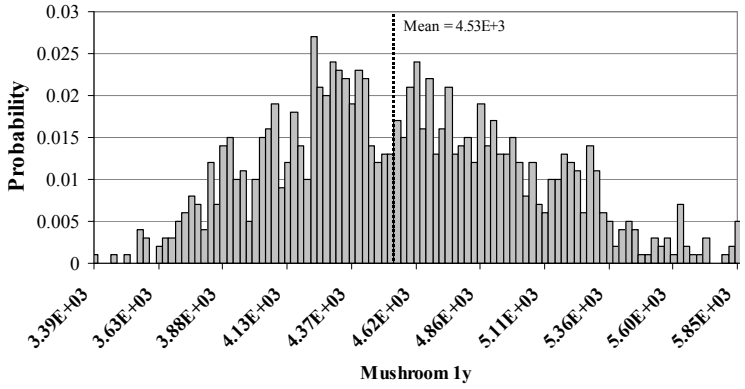
Table 3.25 Mean (μ) and standard deviation (σ) of model inputs for the Tarvisio scenario.

	Bulk Density			Thickness			Interception		Tree roots		Mushroom mycelium			Understorey roots			T_{agg}	TF	
	Litter	Organic Soil	Mineral Soil	Litter	Organic Soil	Mineral Soil	Litter	Tree	Organic soil	Mineral soil	Litter	Organic soil	Mineral soil	Litter	Organic soil	Mineral soil	Mushroom	Understorey	RoeDeer
μ	4.0E+01	1.5E+02	5.0E+02	5.0E-03	2.0E-02	2.4E-01	2.5E-01	7.5E-01	7.0E-01	3.0E-01	3.0E-01	5.0E-01	2.0E-01	3.0E-01	4.0E-01	4.0E-01	3.9E-01	5.4E-02	5.0E-02
σ	8.0E+00	3.0E+01	1.0E+02	1.0E-03	4.0E-03	4.8E-02	5.0E-02	1.5E-01	1.4E-01	6.0E-02	6.0E-02	1.0E-01	4.0E-02	6.0E-02	8.0E-02	8.0E-02	7.9E-02	1.1E-02	1.0E-02

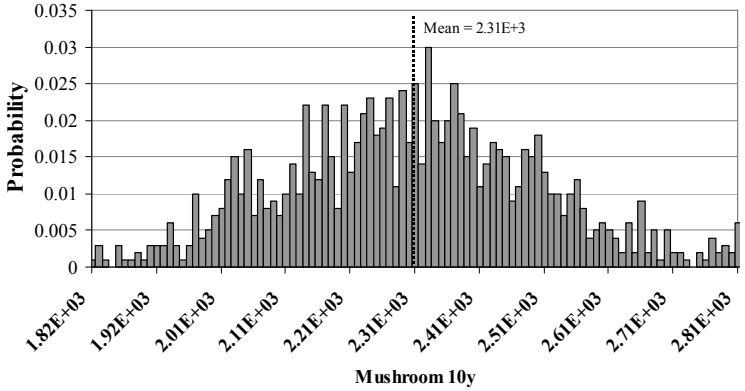
3.3.3.1 Mushroom

The pdfs can be regarded as normal (Figure 3.15a, b and c), coefficient of skewness < 0.5, and the level of uncertainty is rather low, the NG_{75} ranges between 0.26 and 0.16 (Table 3.12) and there are no major differences between the uncertainties evaluated at the three time steps.

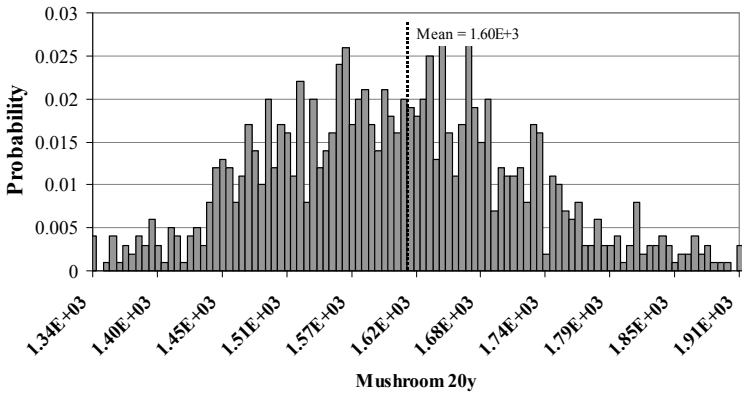
In the short period after ^{137}Cs deposition, the system is strongly influenced by the variation of k_4 parameter (Figure 3.16), which is part of the equation that describes the flux between tree external and soil litter layer. The effect of this parameter is completely negligible in the long-term since the contamination on the tree external is reduced soon after deposition by weathering.



(a)



(b)



(c)

Figure 3.17 Probability density functions describing the ^{137}Cs activity concentration in mushroom at 1, 10 and 20 year after contamination.

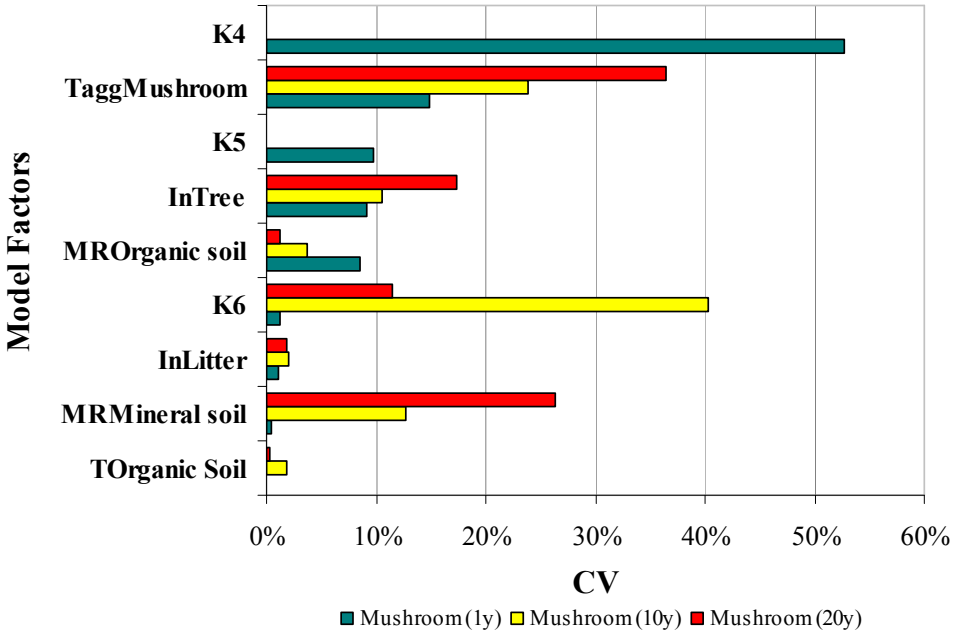


Figure 3.18 Sensitivity analysis on the ^{137}Cs activity concentration in mushroom (Bq/kg).

The influence of the percentage of mushroom mycelium present in the organic (MROrganic soil) and mineral soil layer (MRMineral soil) Figure 3.16), changes with time and highlights the model simulation of ^{137}Cs percolation through soil profile. Consequently in the short and medium term the main part of the soil profile contaminated by radiocaesium is the organic layer, therefore the percentage of mycelium present in this layer represents the major source of contaminant for plants.

On the other hand in the long term ^{137}Cs is assumed to be mostly accumulated in the lower part of the soil profile which is the mineral layer. The ^{137}Cs absorbed by the mycelium present in the mineral layer is the principal contribution to mushroom contamination.

The k_6 parameter is influential only at 10 years after the initial contamination (Figure 3.16). This global parameter is part of the equation used to describe the flux between organic and mineral layer. Therefore at 10 years after contamination, most of the ^{137}Cs has percolated through the soil profile and it is mostly located in the mineral soil layer.

The UA and SA show that the RIFE1 prediction of ^{137}Cs activity concentration in mushroom is characterised by a low level of uncertainty and further more that the constructed model is comparable to the conceptual model. In fact the sensitivity of the model to local and global parameters clearly reflects the simulation of the gradual percolation of ^{137}Cs through soil profile and therefore the importance of considering the ^{137}Cs concentration at each soil layer (i.e. organic and mineral) at different time points.

3.3.3.2 Bark

The uncertainty of the predicted ^{137}Cs activity concentration in bark is rather high at the three time points investigated, NR_{75} ranges between 1.33 and 1.25. In addition, the probability distribution is characterised by a positive skewness, coefficient of skewness > 1.3 (Figure 3.17a, b and c).

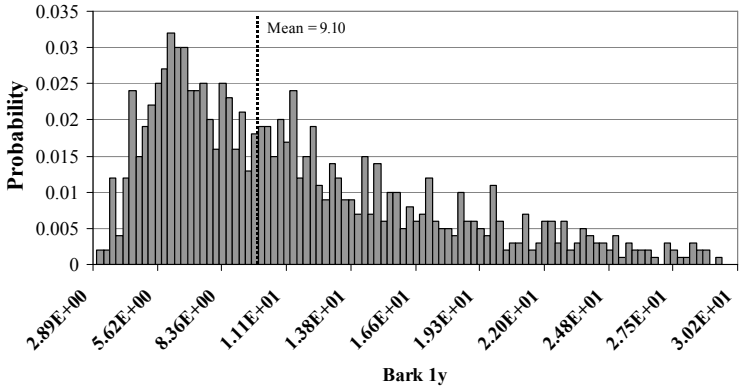
There is no substantial variation in the estimated NR_{75} at the three time points, therefore it is possible to assert that time does not affect the uncertainty on the prediction (Table 3.12).

The dynamics associated with the bark compartment are different from that observed for mushroom. The SA results show that the uncertainty of the predicted ^{137}Cs activity concentration in bark is primarily influenced by the variation of the k_3 parameter (Figure 3.18), which is the global parameter used to evaluate the fluxes between the internal tree compartment and the two soil layers (organic and mineral layers). This is because the main source of tree contamination by ^{137}Cs is not from the foliar or bark absorption but from root uptake from the soil (i.e. organic and mineral layers).

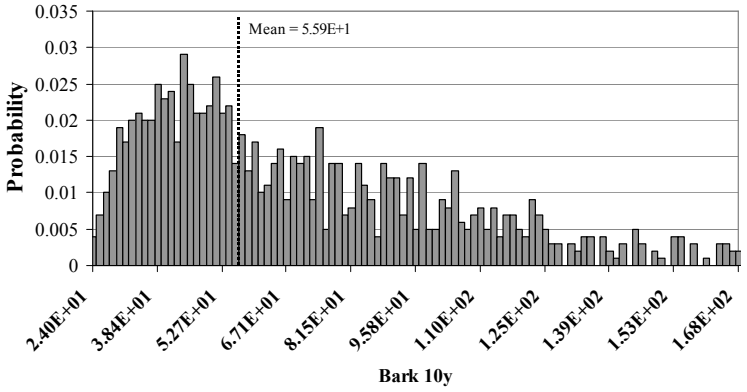
However in the short term, one year after contamination, even the parameters k_4 and k_5 appear to have an effect (Figure 3.18). Those parameters are part of the equation to evaluate the flux between the external parts of the tree to the litter layer and between the litter to the organic soil layer. ^{137}Cs is initially deposited on the tree surfaces and on the litter layer, however within the first year the ^{137}Cs deposited on the tree external is removed by weathering processes and it is re-deposited on the litter layer. With time, it progressively migrates from the soil surface to the organic layer.

On the other hand, in the middle and long term the system shows a minor sensitivity to the k_2 parameter (Figure 3.18). This global parameter is part of the sub-model used to determine the level of contaminant that is transferred from the tree internal to the soil litter layer. The sensitivity of the system to this process ten years after the contamination is in accord with the hypothesis that in the short term the tree is slowly contaminated mainly by root uptake and when such a contamination has reached a significant level the flux between tree and litter layer takes place. However as the level of contaminant transferred is regarded as low, negligible system sensitivity to such a process is expected.

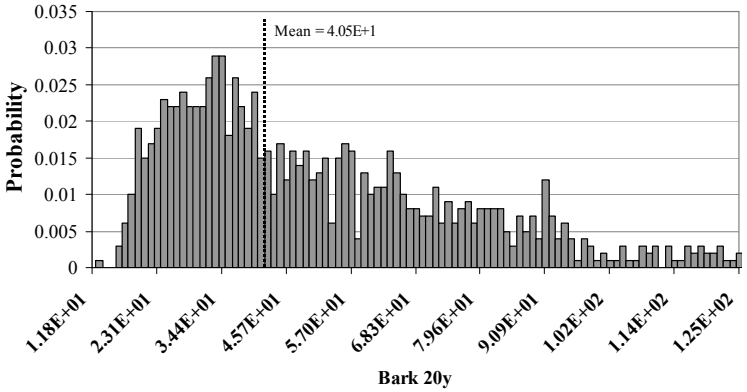
The uncertainty associated with the prediction of ^{137}Cs activity concentration in bark is high (Table 3.10) and the distribution is lognormal, however the SA shows that the system is comparable to the conceptual model since its sensitivity to model factors can be clearly illustrated by the ecosystem dynamics.



(a)



(b)



(c)

Figure 3.19 Probability density functions of the ^{137}Cs activity concentration in bark at the three time periods investigated.

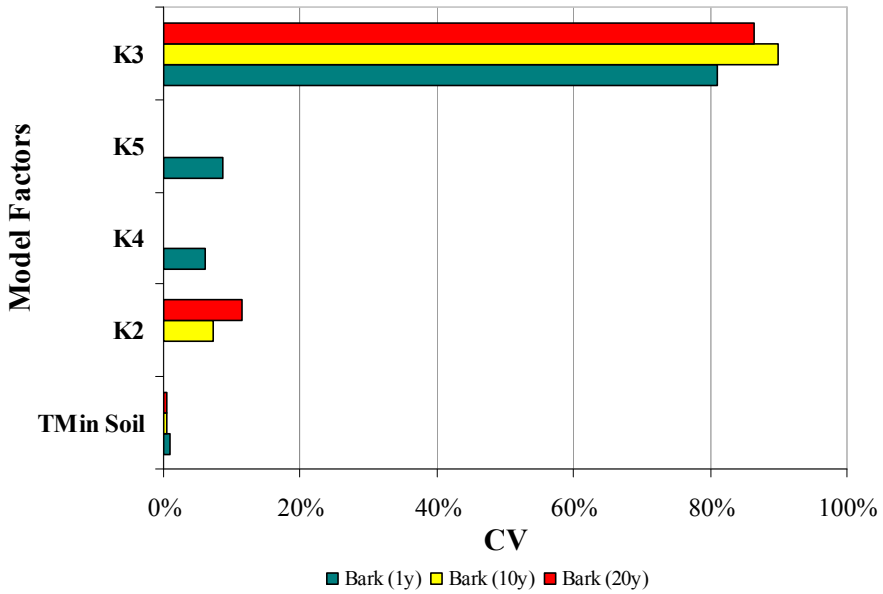


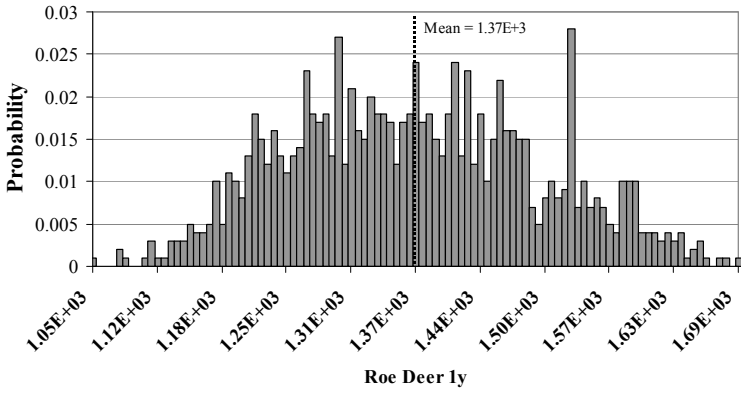
Figure 3.20 Sensitivity analysis results regarding the Bark compartment.

3.3.3.3 Roe deer

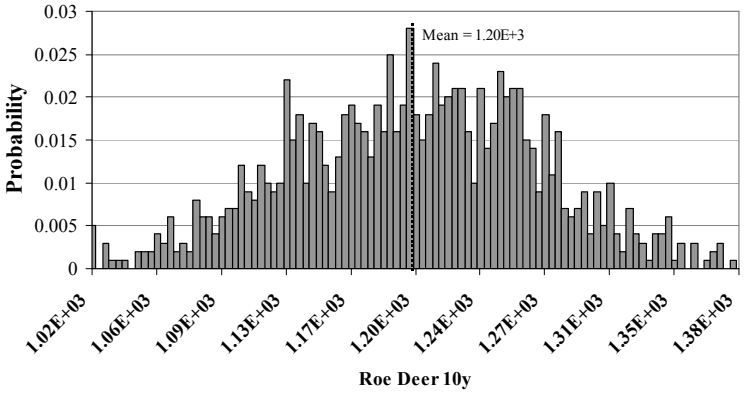
The predicted ^{137}Cs activity concentration in Roe deer (Bq/kg) is characterised by a low level of uncertainty (Figure 3.19a, b and c) with NR_{75} ranging between 0.22 and 0.14.

The NR_{75} at the three time points investigated shows a decrease in uncertainty after the first ten years of contamination, after which an approximate equilibrium seems to be reached (Table 3.11).

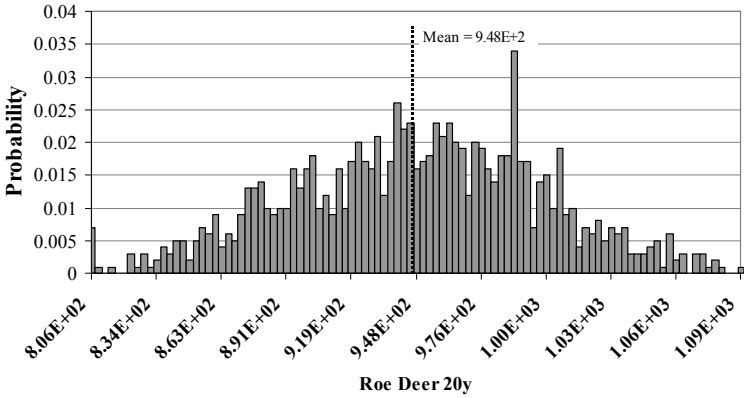
In the first year after the ^{137}Cs deposition, the system is primarily influenced Figure 3.20. The TF value implemented has only a secondary role. A relatively fraction of the ^{137}Cs is located in the tree external surface and a minor proportion is in the litter layer. Consequently, k_4 controls the flux between tree and soil. In the long term, nevertheless, the predicted ^{137}Cs activity concentration in roe deer is very sensitive to any variation in the TF value and to the k_3 parameter (organic/mineral soil layer – tree internal transfer coefficient). This is because in the long term ^{137}Cs is accumulated in the soil layers (mainly organic and mineral) and the contamination of roe deer is therefore more sensitive to the uncertainty on the TF value adopted, (Figure 3.20).



(a)



(b)



(c)

Figure 3.21 Probability density functions of the roe deer contamination prediction (Bq/kg) at 1, 10 and 20 years after contamination.

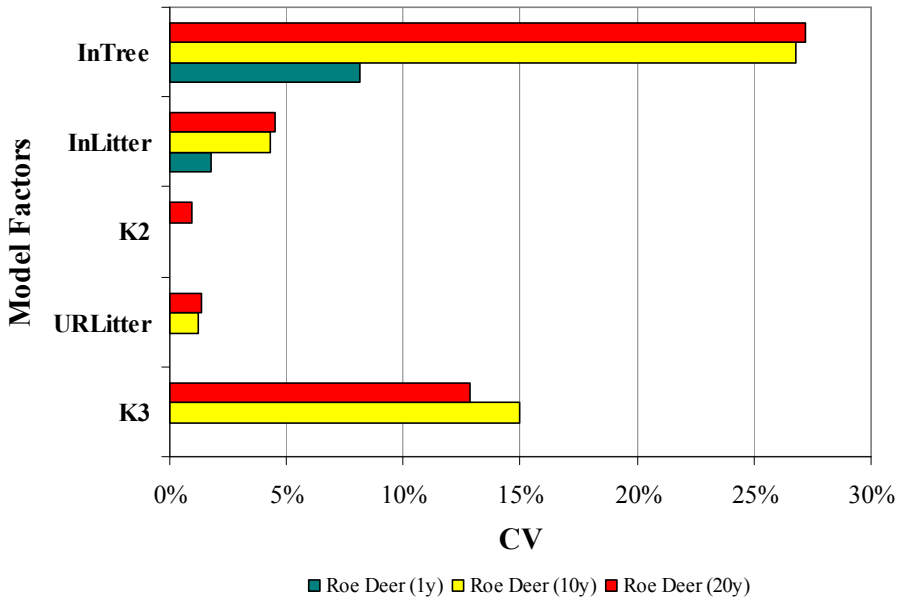


Figure 3.22 Results of the sensitivity analysis performed on the Roe Deer predictions.

3.3.3.4 Conclusions

The uncertainty and sensitivity have been evaluated for three model outputs: mushroom (Bq/kg), bark (Bq/kg) and roe deer (Bq/kg). In general the model is characterised by a low level of uncertainty, NR_{75} is lower than 0.3 for mushroom and roe deer and 1.3 for bark (Table 3.11). In addition time does not appear to have any major effect on the estimated uncertainty.

The results of the UA and SA demonstrate that RIFE1 is substantially similar to the conceptual model. The sensitivity to optimised local parameters should not be considered a limitation since RIFE1 has been designed as a local model.

3.3.4 FORM uncertainty and sensitivity analysis

The TEMAS semi-natural model (FORM) has been tested for uncertainty and sensitivity on three model outputs: ^{137}Cs activity concentration in bark, mushroom and roe deer. Similar experiments to RIFE1 have been performed:

1. Determining the uncertainty and model sensitivity on the three end-products, considering a generic scenario for the local parameters (Table 3.13).

2. Comparing the uncertainties and sensitivities at different time steps, 1, 10 and 20 years after initial contamination in order to establish the potential influence of time on the output uncertainty (Table 3.14).

The correlation between model factors has not been included in the MCS as insufficient data was available to estimate *r*.

Table 3.26 FORM generic scenario: OG (*) and local parameters ($^+$).

	μ	σ
R_{12}^*	1.00E-02	2.00E-03
R_{23}^*	1.40E+00	2.80E-01
R_{34}^*	3.50E-01	7.00E-02
R_{45}^*	3.50E-02	7.00E-03
R_{55}^*	1.00E-02	2.00E-03
$TF_{mushroom.}^*$	1.00E-01	2.00E-02
$TF_{roedeer}^*$	4.75E-02	9.50E-03
Forest age at contamination (y)$^+$	1.00E+01	2.00E+00
Contamination – harvest (y)$^+$	7.00E+01	1.40E+01

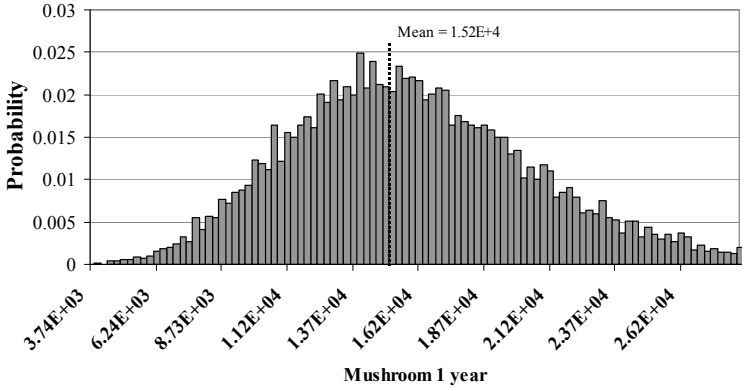
Table 3.27 NR_{75} for the three model outputs investigated.

	NR_{75}		
	1y	10y	20y
Mushroom	7.08E-01	4.68E-01	4.64E-01
Bark	4.72E-03	4.70E-02	9.48E-02
Roe Deer	2.51E-01	1.57E-01	1.57E-01

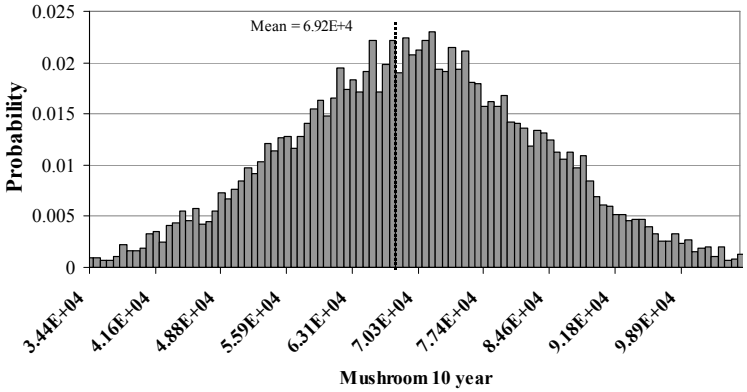
3.3.4.1 Mushroom

The predicted ^{137}Cs activity concentration in mushroom is normally distributed (Figure 3.21a, b and c) and the uncertainty is rather low, $NR_{75} > 0.7$ (Table 3.14). The variation of NR_{75} with time is negligible, however a reduction is

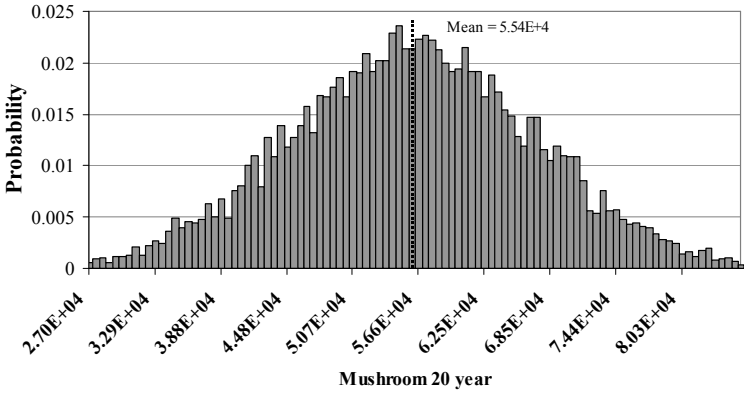
observable within the first ten years after contamination while a quasi-constant NR_{75} is recorded for the following ten-year period (Table 3.14).



(a)



(b)



(c)

Figure 3.23 Probability density functions for predicted ^{137}Cs activity concentration in mushroom (Bq/kg).

The results of the SA (Figure 3.22) show that the main factor influencing the output uncertainty is the TF value adopted. However in the short term the system is also moderately sensitive to the R_{34} and R_{23} factors, which represent the fluxes between litter to organic layer and leaves to litter respectively.

The prediction of ^{137}Cs activity concentration in mushroom is performed using a TF approach. This explains the strong sensitivity of the system to the TF value adopted, which evaluates the contamination of mushroom as function of the total ^{137}Cs activity concentration in the organic and mineral soil layers.

In the first year of contamination, the radionuclide migration between litter and organic soil layer (R_{34}) and the foliar weathering (R_{23}) are important processes, since all the ^{137}Cs present in the ecosystem is located on soil and plant surfaces. Nevertheless, in the long-term the ^{137}Cs present in the litter layer has migrated to the lower soil layer and the foliar weathering and senescence do not represent a source of contamination. Consequently, the main source of uncertainty can be attributed to the TF value adopted.

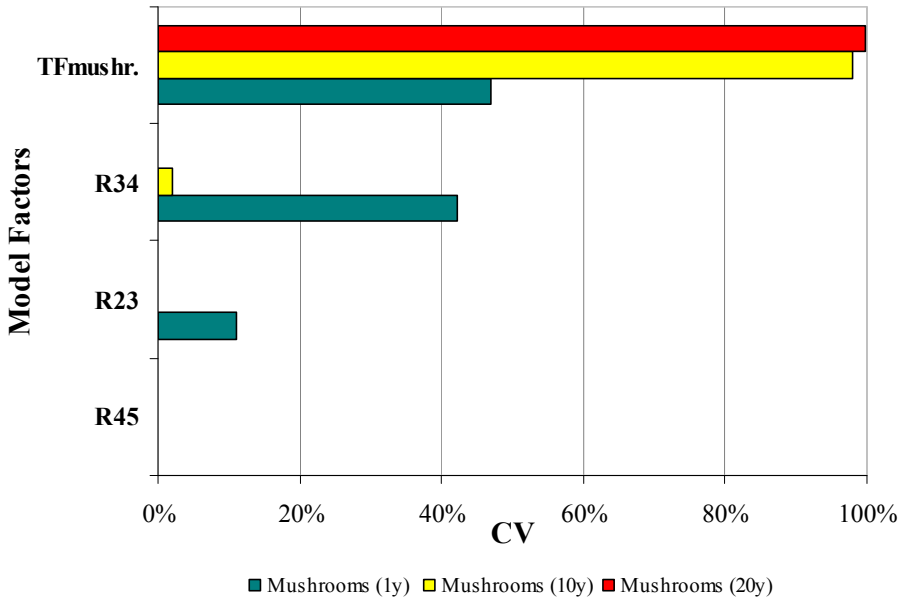


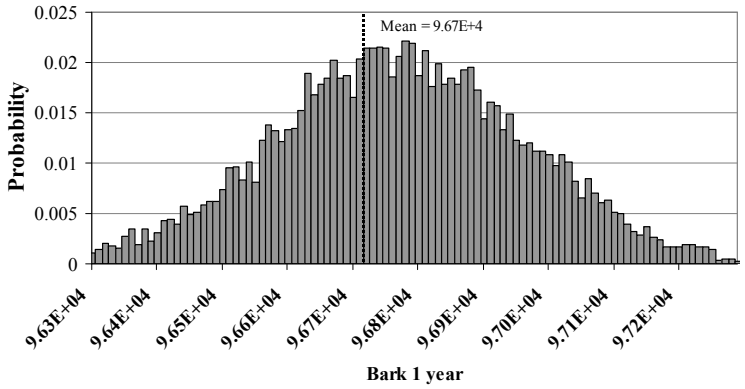
Figure 3.24 Results of the sensitivity analysis performed on the predicted mushroom contamination.

The prediction of mushroom contamination is characterised by a low level of uncertainty, however the sensitivity analysis shows that such an uncertainty is mainly function of the TF value adopted. This might represent a limitation in the use of this system, since the TF values are extremely variable between sites and even within the same site.

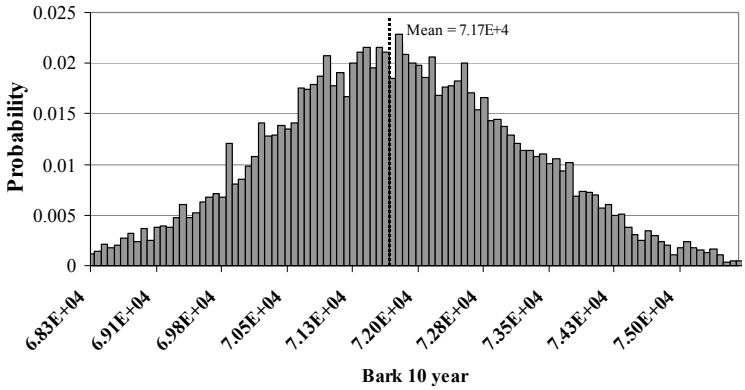
3.3.4.2 Bark

The bark probability distributions are normally distributed (Figure 3.30a, b and c), coefficient of skewness > 0.02, and the levels of uncertainty are low, NR_{75} ranges between $4.73*10^{-3}$ and $9.42*10^{-2}$, (Table 3.14).

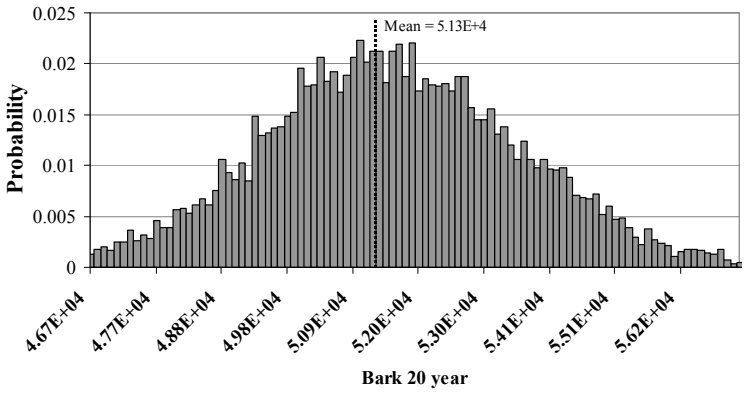
The trend of NR_{75} as a function of time shows a gradual increase with time. However the uncertainty at twenty years after contamination, which is the highest recorded, is still relatively low, therefore it should not have any major effect on the model reliability.



(a)



(b)



(c)

Figure 3.25 Probability density functions of the predicted bark contamination (Bq/kg).

The model is mainly sensitive to the model parameter R_{12}, CV is greater than 90% (Figure 3.24), which is part of the sub-model used to evaluate the flux between bark and litter layer.

The strong influence of the R_{12} factor on the prediction uncertainty is due to the model design, in fact the ^{137}Cs activity concentration in bark is estimated as a function of the initial deposition and it is reduced by bark senescence. Although the constructed model is in agreement with the conceptual model, the model prediction is questionable since it is entirely based on a global parameter and it does not take into consideration the process of bark formation, where the ^{137}Cs present in the wood is partially transferred into the bark.

3.3.4.3 Roe deer

The prediction of Roe deer contamination is characterised by a low level of uncertainty, NR_{75} ranges within 0.25 and 0.16 (Table 3.14) and the pdfs can be considered as normally distributed since the coefficient of skewness are lower than 0.5, (Figure 3.25a, b and c).

The predicted activity concentration in roe deer is estimated using the same approach adopted for the mushroom contamination, which is the Transfer Factor approach and the animal contamination is a function of the ^{137}Cs activity concentration in the organic and mineral soil layers.

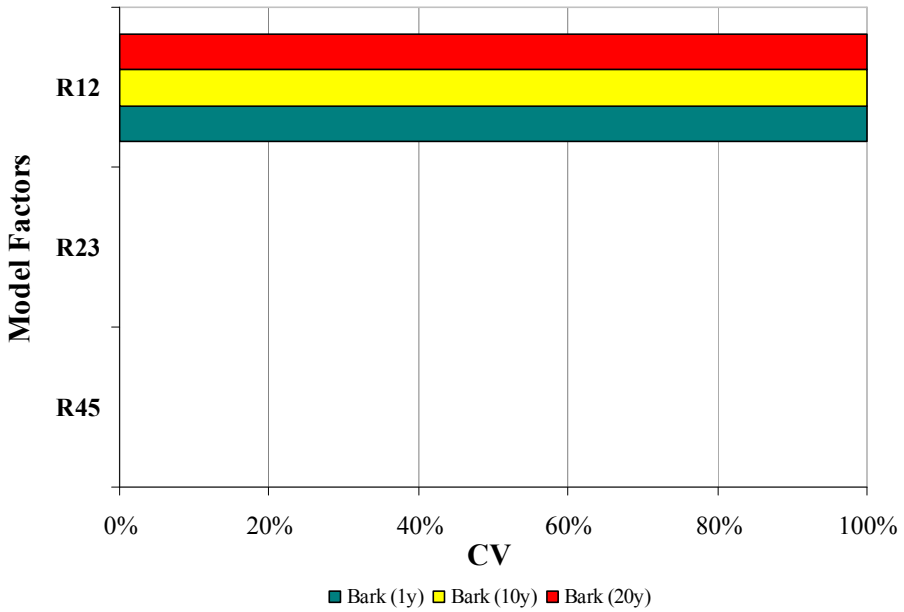
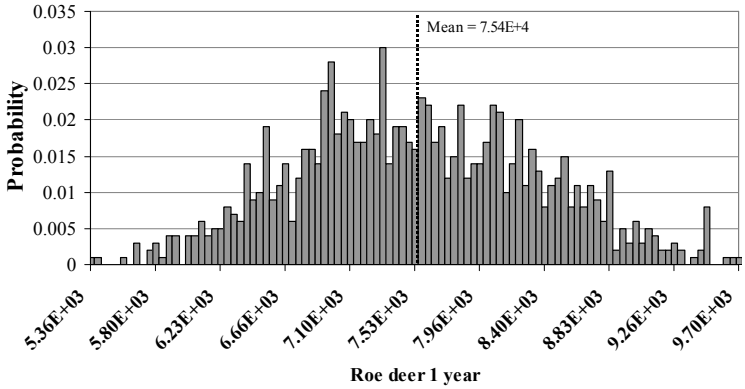
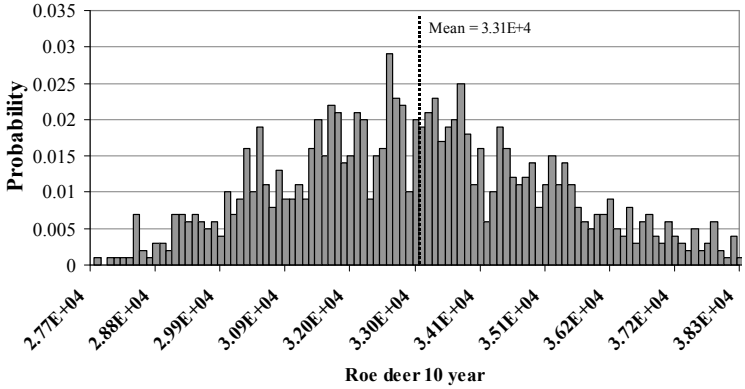


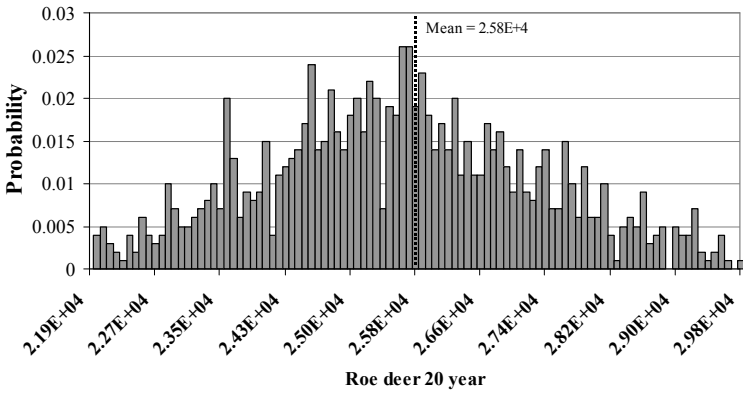
Figure 3.26 Sensitivity analysis performed on the predicted ^{137}Cs activity concentration in bark.



(a)



(b)



(c)

Figure 3.27 Probability density functions of the estimated ^{137}Cs activity concentration in roe deer meet (Bq/kg).

The similarities between the methods used to estimate the mushroom and roe deer contamination are reflected in the SA results (Figure 3.26). In the first year the system is sensitive to R_{34} (litter-organic soil layer transfer coefficient) and R_{23} (leaves-litter transfer coefficient) since the ^{137}Cs has deposited on the plants and soil surface, therefore the leaves senescence and ^{137}Cs migration from litter to organic layer represent significant processes. However, in the long term, 10 and 20 years after contamination, the only factor to which the system is sensitive is the transfer factor adopted, as in the long term a significant fraction of the total ^{137}Cs is located in the organic and mineral layer.

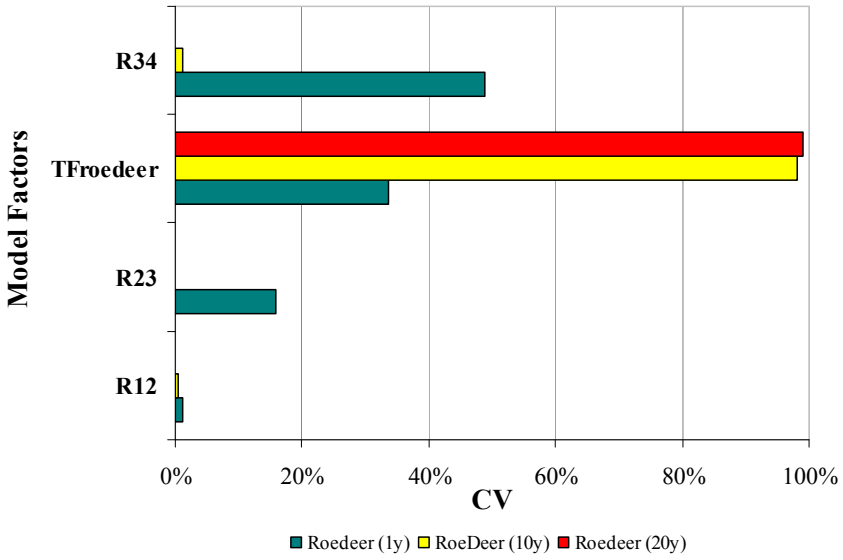


Figure 3.28 Results of the sensitivity analysis performed on the Roe Deer compartment.

The strong influence of the TF value on the model uncertainty represents a limitation in the use of this model since the TF values, which describe the transfer of ^{137}Cs between the soil to animals are extremely variable.

3.3.4.4 Conclusions

FORM has been designed adopting a dynamic compartment approach and its design is rather similar to RIFE1. The three outputs investigated are characterised by a low uncertainty, $NR_{75} > 0.3$, Table 3.14.

However considering the overall results from the SA and UA, FORM has not performed as well as RIFE1. FORM is mostly sensitive to the TF used; therefore the model prediction accuracy is entirely based on the reliability of the TF values adopted. Therefore FORM can be considered only partially in agreement with the conceptual model, since the model sensitivity to the TF values used overshadows the modellers' attempt to create a system which simulates ecosystem processes such as ^{137}Cs percolation through the soil profile and plant contamination through root uptake. As a result FORM could be largely simplified and the resulting model would not differ from a simple TF model.

3.4 CONCLUSIONS

The uncertainty and sensitivity analyses have highlighted differences between conceptual and constructed models regarding the importance of input data and robustness of the model design adopted. The following specific conclusions can be drawn:

- The model implemented in SAVE for the rural environment (Absalom et al, 2001) has showed that the predicted CF is characterised by a large uncertainty (NR_{75} ranging between 4.5 and 6.3) which is not affected by time. In addition, the consideration of the correlation between model factors (OM-pH and k_1-k_2) does not have any relevant effect on the estimation of the uncertainty.

  The model uncertainty is entirely dominated by the k_2 parameter, which can be considered the source of the model uncertainty. The sensitivity analysis results obtained from the "correlation" experiment suggest that the CF is highly sensitive

not only to the k_2 but also to the k_1 parameter, CV equal to 50% and 45% respectively. However, this finding does not modify what was previously concluded. The sub-model used to estimate CF accounts for the large uncertainty on the model output.

Further investigation has been performed on the CF sub-model and the results are presented in chapter 4.

- The TEMAS rural model is characterised by a model design which reduces the amount of input data that the user has to supply, since a database provides the model with soil and scenario characteristics. The uncertainty analysis on three model outputs (soil, edible and inedible part of the plant, Bq/kg) reveals that the model outputs are characterised by a low level of uncertainty, $NG_{75} > 0.6$. However, the model is particularly sensitive to the global parameters which are part of the database. This represents a major weakness of the system as the accuracy of model predictions is entirely based on the quality of the database. The model may only provide an adequate prediction if the scenario data included in the database corresponds to the characteristics of the scenario considered. The database has to be carefully assessed, since it can strongly affect the final model prediction and its implementation.

  In addition, a test on the model design robustness has suggested that the database approach does not show the flexibility which would be required for a decision support tool.

- Three model outputs of RIFE1 have been tested for uncertainty and sensitivity: Mushroom (Bq/kg), Bark (Bq/kg) and Roe Deer (Bq/kg). The estimated probability distributions are normal and the estimated NR_{75} values indicate that mushroom and roe deer are characterised by a low level of uncertainty $NR_{75} > 0.3$, while the predicted activity concentration in bark a higher uncertainty, NR_{75} ranges between 1.24 and 1.33.

  The results of the sensitivity and uncertainty analysis illustrate the similarities between conceptual and constructed model.

- FORM is the forest model implemented in TEMAS and its design is similar to the one used to develop RIFE1. The three outputs investigated, Mushroom

(Bq/kg), Bark (Bq/kg) and Roe Deer (Bq/kg), are characterised by a low uncertainty, $NR_{75} > 0.3$. However, unlike RIFE1, the model is mostly sensitive to the transfer factors (TF) used, therefore the model prediction accuracy may be entirely due to the reliability of the TF value adopted. In addition the model cannot be considered entirely comparable to the conceptual model.

4. INCREASING THE RESEMBLANCE BETWEEN CONCEPTUAL AND CONSTRUCTED MODELS: A CASE STUDY

The uncertainty (UA) and sensitivity analysis (SA) discussed in chapter 3 determined that the model developed by Absalom et al (2001) has little resemblance with the conceptual model. The main objective of Absalom et al was to develop a semi-mechanistic model which could predict the soil-to-plant transfer of radiocaesium based on soil characteristics, i.e. pH, exchangeable potassium, organic matter and clay content. However, the model is not significantly sensitive to input parameters. The SA results show that the model is highly sensitive to the k_2 parameter (CV > 90%), which is part of the CF–m_k relationship. While the k_9 parameter (RIP_{clay} sub-model) was identified as the second most important factor, contributing 3 – 4% of the prediction uncertainty.

The model sensitivity is entirely dominated by optimised global parameters (OG), which are regarded as the main cause of the high level of uncertainty affecting the model prediction, i.e. TF_{min}, TF_{med} and TF_{max}, NR_{75} ranges between 4.5 and 6.8.

Consequently, the aim of this chapter is to re-evaluate the Absalom model in order to increase its degree of similarity with the conceptual model.

This chapter will be divided into three sections:

- *Introduction*, where a detailed overview of the Absalom model (Absalom et al, 2001) will be provided,
- *Revised model description*, where the re-evaluation of two sub-models will be discussed,
- *Results and discussion*, where the predictions of the revised and the Absalom model are compared on three aspects: uncertainty, sensitivity and degree of agreement with observations.

4.1 INTRODUCTION

4.1.1 Absalom model description

The model developed by Absalom et al (2001) is an improved version of a model published previously by the same group (Absalom et al, 1999). The 1999 model version had the limitation that it could only be used for mineral soils, as the labile Cs distribution coefficient k_d (dm^3/kg) was considered only as a function of the clay and exchangeable K^+ (mol/l) content in soil. The later model (Absalom et al, 2001) addressed this aspect and consequently it can be considered as a more generalised model, which can potentially account for the role of organic matter in radiocaesium availability and thus subsequent uptake by plants.

The model parameterisation was performed using two datasets, Smolders et al (1997) and Sanchez et al (2000), while the accuracy of the model prediction, i.e. TF, was tested comparing the model output with empirical data derived from Nisbet et al (1999). The Absalom model parameters are presented in table 4.1.

Table 4.1. Values of fitted and constants (*) used in the Absalom model (Absalom et al, 2001); Each value is reported with the standard error (±).

Model parameters	Value	Equations
k_1	2.57 ± 0.1	3.3
k_2	1.56 ± 0.32	3.3
k_3^*	3.37	3.5
k_4^*	0.16	3.5
k_5^*	-34.66	3.6
k_6^*	29.72	3.6
k_7	4.18 ± 0.27	3.9
k_8	0.043 ± 0.017	3.10
k_9	1.74 ± 0.17	3.10
k_G^{clay}	3.18 ± 0.44	3.7
k_G^{humus}	2.32 ± 0.1	3.7
k_{fast}^* (1/d)	$1.9*10^{-3}$	3.12
k_{slow}^* (1/d)	$1.9*10^{-4}$	3.12
P_{fast}^*	0.814	3.12
CEC^{clay*} ($cmol_c/kg$)	50	3.4

The model (Absalom et al, 2001) can be divided into four main parts: m_k, k_{dl}, D_f and CF (Figure 4.1). A detailed overview of each part is given in the subsequent sections.

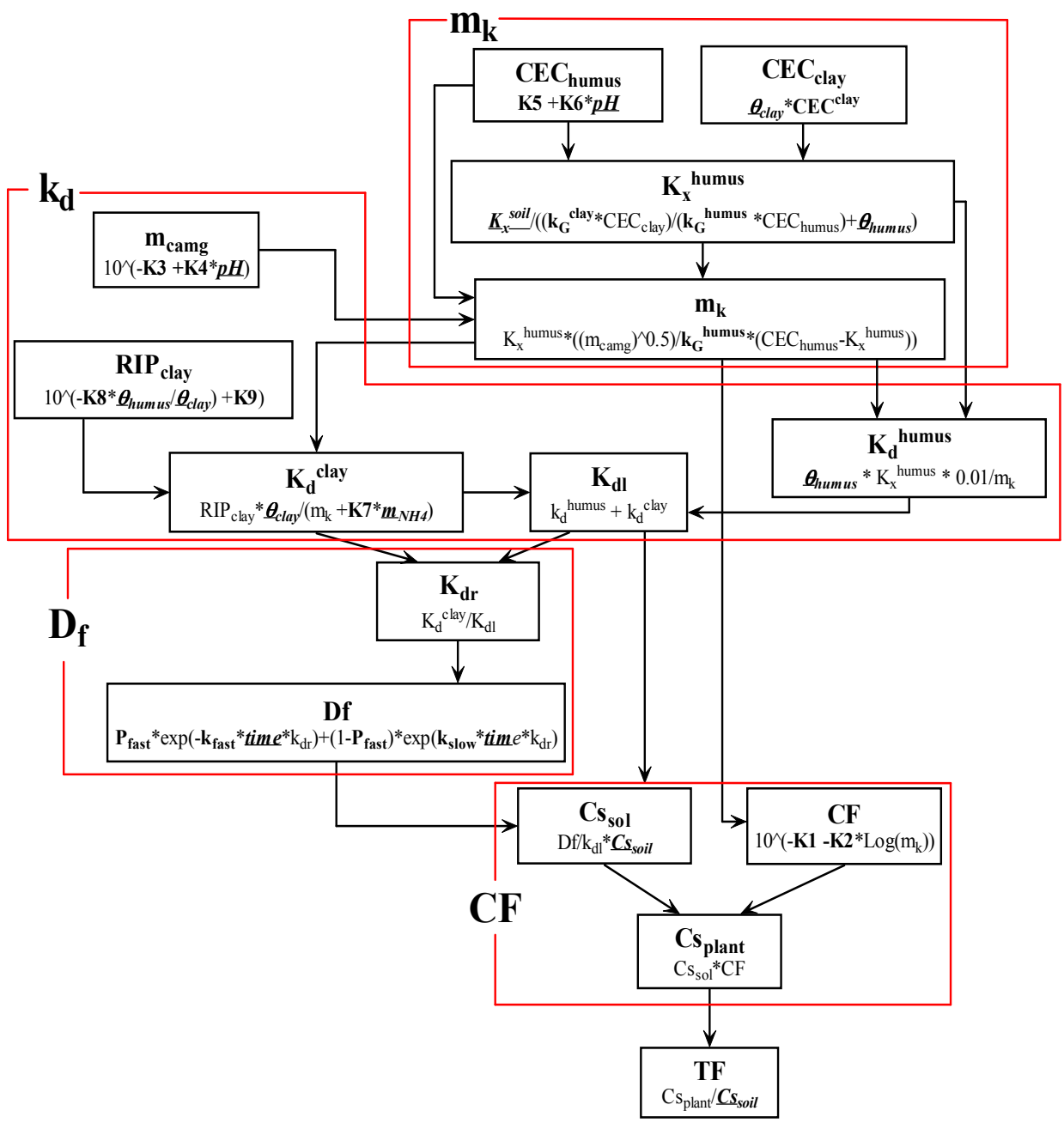


Figure 4.1 Four parts comprise the Absalom model: m_k, k_d, Df and CF. The m_k element estimate the K in soil solution, the k_d element, the Df element introduces the time dependency into the system and finally the CF element estimates the radiocaesium activity concentration in plants. The italic underlined factors are local parameters while the global optimised parameters are reported as bold.

4.1.1.1 CF sub-model

The CF sub-model estimates the plant contamination due to Cs uptake (Cs_{plant}, Bq/kg) as the product between CF and the estimated Cs in soil solution (Cs_{sol}, Bq/dm^3), equation 4.1.

$$Cs_{plant} = CF * Cs_{sol} \tag{4.1}$$

where

CF is the soil-to-plant Concentration Factor.

Cs_{sol} is the Cs in soil solution (Bq kg^{-1}).

The application of equation 4.1 requires the estimation of Cs in soil solution (equation 4.2) and the soil-to-plant CF (equation 4.3).

$$Cs_{sol} = \frac{D_f}{k_{dl}} Cs_{soil} \tag{4.2}$$

where

D_f is a factor that accounts for the change of labile Cs with time (equation 4.12).

k_{dl} is the estimated labile-fixed Cs distribution coefficient (dm^3 kg^{-1}) (equation 4.8).

Cs_{soil} is the total radiocaesium present in soil (Bq kg^{-1}).

A study by Smolders et al (1997) demonstrated that plant uptake of radiocaesium, described as soil-to-plant CF, is negatively correlated to the K^+ concentration in soil solution (mmol/l), Figure 4.2. These results have been supported by several other independent studies (e.g. Sanchez et al, 1999), therefore a linear model was implemented, equation 4.3.

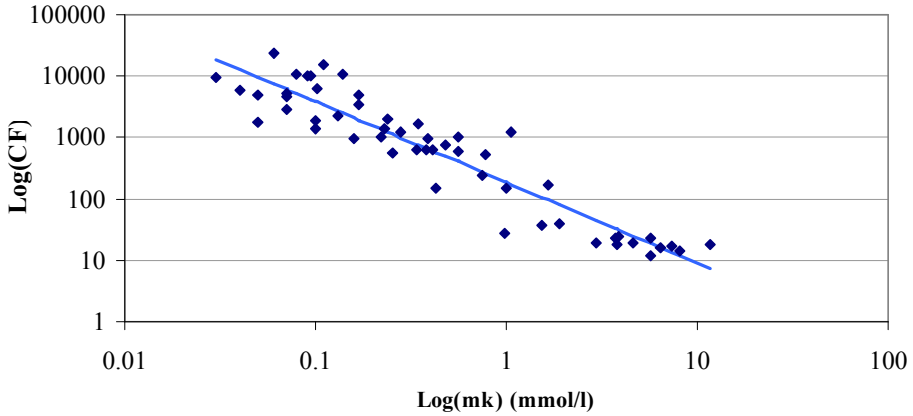


Figure 4.2 The soil-to-plant concentration factor (CF) is negatively correlated to potassium in soil solution (m_k, mmol/l). The solid line (—) represents the CF-m_k linear model, while the points (♦) represent the measurements derived from the Smolders et al (1997) and Sanchez et al (1999) work.

$$Log\left(CF\right) = -k_1 - k_2 Log\left(m_k\right) \tag{4.3}$$

where

k_1 and k_2 are model parameters which have been estimated by fitting the model to observed data.

m_k is the K^+ concentration in soil solution (mmol/l).

4.1.1.2 m_k sub-model: K^+ concentration in soil solution.

The concentration of K^+ in soil solution (m_k) is estimated by equation 4.4, which has been developed assuming that the soil solution predominantly comprises Ca^{2+} and Mg^{2+} ions, which therefore can be regarded as the main competitors for absorption of the free potassium ions onto the exchange sites of soil particles. Given this, the sum of the calcium and magnesium cation concentrations (m_{Ca+Mg}) can realistically be approximated by the difference between the soil cation exchange capacity (CEC, $cmol_c\ kg^{-1}$) and the exchangeable K^+ (K_x^{sol}, $cmol_c\ kg^{-1}$).

Radiocaesium is assumed to interact primarily with two exchangeable surfaces, clay and organic matter. Their relative selectivity, for a pair of cations, is described by the Gapon exchange coefficient (k_G) (Barber, 1984). As a result, m_k is estimated in terms of equilibrium between K^+ absorbed by organic matter and clay.

$$\left[m_k\right] = \frac{K_x^{humus}\sqrt{\left[m_{Ca+Mg}\right]}}{K_G^{humus}\left(CEC^{humus} - K_x^{humus}\right)} \tag{4.4}$$

where

K_x^{humus} and K_x^{clay} are the exchangeable K^+ ($cmol_c\ kg^{-1}$) on the organic and clay surfaces respectively (equation 4.7).

m_{Ca+Mg} is the combined concentration of calcium and magnesium cations in solution (mol/l) (equation 4.5)

k_G^{humus} and k_G^{clay} are the Gapon exchange coefficients for the clay and organic fractions estimated fitting the model to observed data (Table 4.1).

CEC^{humus} and CEC^{clay} are the cation exchange capacities of the organic matter and the clay surfaces (equation 4.6 and Table 4.1).

The (Ca^{2+} + Mg^{2+}) concentration in soil solution (m_{Ca+Mg}) is estimated using equation 4.5, which assumes a positive correlation between m_{Ca+Mg} and pH, and

$$Log\left[m_{Ca+Mg}\right] = -k_3 + k_4 pH \tag{4.5}$$

where

k_3 and k_4 are constants, for (Table 4.1) derived assuming m_{Ca+Mg} varies between 0.001 and 0.009 moles/l as pH ranges between 2 and 8.

The study of Hellings and co-workers (Hellings at al, 1964) aimed to establish the effects of pH on the clay and OM contribution on the soil CEC. Soil characteristics of samples from sixty soils were determined: pH ranged between 2.5 and 8 and the relative CEC for organic matter (CEC^{humus}) was observed to vary between 36 $cmol_c$/kg and 213 $cmol_c$/kg respectively. A significant linear relationship was formed, equation 4.6.

$$CEC^{humus} = k_5 + k_6 pH \tag{4.6}$$

where

k_5 and k_6 are constants and the values are provided in table 4.1.

The CEC for clay (CEC^{clay}) has been observed to range between 38 $cmol_c/kg$ (pH 2.5) and 64 $cmol_c/kg$ (pH 8) with an average of 53 $cmol_c/kg$, which supports the constant value adopted by Absalom et al (1999) for CEC^{clay} of 50 $cmol_c/kg$.

A key element of the m_k sub-model is the adsorbed exchangeable organic matter potassium (K_x^{humus}), equation 4.7, which is evaluated assuming that the total exchangeable K^+ has a minor role on the total CEC.

$$K_x^{humus} = \frac{K_x^{soil}}{\left(\left(k_G^{clay} * CEC^{clay} * \theta_{clay} \right) \Big/ \left(k_G^{humus} * CEC^{humus} \right) \right) + \theta_{humus}} \qquad (4.7)$$

where

θ_{clay} and θ_{humus} are the measured gravimetric clay and humus contents ($g\ g^{-1}$)

k_G^{clay}, k_G^{humus} and CEC_{clay} values are reported in the table 4.1

4.1.1.3 k_{dl} sub-model. Labile Cs distribution coefficient

The equilibrium between free Cs ions in soil solution and radiocaesium fixed on clay minerals or labile adsorbed by organic fraction is described by the labile Cs distribution coefficient (k_{dl}, dm^3/kg), equation 4.8.

The k_{dl} of the whole soil is considered a function of both the k_d for organic matter (k_d^{humus}) and clay (k_d^{clay}). This approach ensures that the sorption by the mineral and organic soil fractions are considered.

$$k_{dl} = k_d^{clay} + k_d^{humus} \qquad (4.8)$$

The labile Cs distribution coefficient for the soil mineral fraction is assumed to be negatively correlated to the potassium and ammonium concentrations in soil solution (m_k and m_{NH4}, respectively), and proportional to the radiocaesium interception potential of clay particles (RIP^{clay}) and gravimetric clay content (θ^{clay}, $g\ g^{-1}$), equation 4.9.

$$k_d^{clay} = \frac{RIP^{clay}\theta^{clay}}{\left[m_k\right] + \left(k_7\left[m_{NH_4}\right]\right)} \quad (4.9)$$

where

m_k and m_{NH4} are the potassium and ammonium in soil solution (mmol/l)

RIP^{clay} is the radiocaesium interception potential of clay particles (mmol/kg)

θ^{clay} is the gravimetric clay content (g g^{-1})

Potassium and ammonium cations are regarded as the main competitors for Cs cations to be specifically adsorbed onto clay particles. Although the effect of NH_4^+ is likely to be a main factor in aerobic soils, it can be ignored in most cases. Nevertheless, in Sanchez et al (2000) data, a significant concentration of NH_4^+ was measured. Therefore, a modification was made to the original 1999 version of the model (Absalom et al, 1999) to include the effect of NH_4^+.

The RIP of a soil can be defined as the capacity of a soil to specifically adsorb the Cs in soil solution and it is an essential element to estimate the radiocaesium soil solution equilibrium. The specific Cs absorption is considered to occur mainly on the clay particle surfaces, consequently the model uses the RIP of the clay fraction (RIP_{clay}, equation 4.10) and does not take into account the organic matter since it will always have a minor role in comparison to the clay effect.

$$Log\left(RIP^{clay}\right) = -k_8\left(\theta_{humus} \Big/ \theta_{clay}\right) + k_9 \quad (4.10)$$

where

k_8 and k_9 are fitted parameters

θ_{humus} and θ_{clay} are the gravimetric organic matter and clay, respectively (g g^{-1})

The RIP_{clay} sub-model has been developed using the Sanchez et al (2000) data, where RIP_{clay} has been observed to decrease to some extent as organic matter increases.

Potassium and radiocaesium free ions are assumed to be subject to an exchangeable absorption on the organic matter (i.e. labile). Therefore, the k_d^{humus} has

been estimated as the ratio between the exchangeable K^+ non-specifically absorbed on the organic fraction (K_x^{humus}, $cmol_c/kg$) and K^+ in soil solution (m_k, mmol/l), equation 4.11.

$$k_d^{humus} = \frac{\left(\theta^{clay} K_x^{humus} 0.001\right)}{\left[m_k\right]} \tag{4.11}$$

Where θ_{clay} is the gravimetric clay (g/g), K_x^{humus} is the exchangeable K^+ non-specifically absorbed on the organic fraction, m_k is the potassium in soil solution and 0.001 is a factor for the conversion of $cmol_c/kg$ to mol/kg.

4.1.1.4 D_f sub-model. The effect of time on the ^{137}Cs bioavailability.

The D_f sub-model introduces the time dependency; the previous sub-models described processes regarding the Cs bioavailability only as a function of chemical soil characteristics.

A dynamic factor (D_f, ranging between 0 and 1), equation 4.12, has been implemented to account for the progressive reduction of radiocaesium in soil solution due to fixation.

$$D_f = P_{fast} e^{\left(-k_{fast} * k_{dr} * t\right)} + \left(1 - P_{fast}\right) e^{\left(-k_{slow} * k_{dr} * t\right)} \tag{4.12}$$

Radiocaesium present in soil solution is considered to be divided into two fractions P_{fast} and (1- P_{fast}), which represent the fast and slow Cs fixation, subject to k_{fast} and k_{slow} fixation rates. Absalom et al (1996) derived a value for P_{fast} of 0.81, for k_{fast} equal to $1.9*10^{-3}$ and k_{slow} of $1.9*10^{-4}$ (1/d) by fitting several literature data sets.

Since it is assumed that the Cs is not specifically absorbed by the organic fraction, the equation 4.12 is corrected by a factor k_{dr}, which is the k_d^{clay} and k_d ratio. Therefore, the lower the estimated k_d for the clay fraction, the slower is the radiocaesium fixation.

4.1.2 Experimental data

Three datasets were used for the development of the revised model presented in this work (i.e. model optimisation, and validation): Smolders (Smolders et al, 1997), Sanchez (Sanchez et al, 1999) and the NRPB (Nisbet et al, 1999) (Table 4.2).

Smolders et al (1997) and Sanchez et al (1999) undertook two independent studies to establish a correlation between plant uptake, expressed in terms of Concentration Factor (CF), and the potassium concentration in soil solution (m_k).

For the purpose of this work, the data of these two studies have been combined into a single dataset, as both studies have been performed on grasses, *Lolium perenne* and *Agrostis capillaries* in the case of Smolders et al (1997) and Sanchez et al (1999) respectively. In addition, they adopted the same experimental methodology.

Smolders sampled thirty grassland topsoils (0-10 cm) in Belgium in 1995. The soils were chosen in order to cover a wide range of soil type: luvisol, fluvisol, podzol, anthrosol, regosol, cambisol, gleysol and histosol. Soil characteristics such as: pH, exchangeable cation content (K, Ca and Mg) and cation exchange capacity (CEC) were determined and the soil RIP was measured (Table 4.2).

Three pot replicates were used for each soil sample and they were spiked with ^{137}Cs and then manually mixed in order to obtain a homogeneous contamination. Ryegrass (*Lolium perenne*) was planted and analysed after 47 - 67 days. This study demonstrated that CF is significantly correlated to K^+ in soil solution (m_k), r^2 0.87.

The study performed by Sanchez et al (1999) had similar objectives, although the emphasis was on organic rich soils. Twenty-three soil samples were collected in north and mid-Wales, Northwest England and Scotland between 1996 and 1997, which included histosols (peats), dystric gleysols and brown soils. The pots were spiked with ^{134}Cs and bent-grass seeds (*Agrostis capillaries*) were planted and analysed after 47 -73 days (Table 4.2).

The NRPB database compiled by Nisbet et al (1999) using published and unpublished data is regarded as a comprehensive set of transfer factors (TFs) and related soil characteristics for 28 soil-crop combinations, covering four soil types and seven crop groups. This database has been used to verify the revisited model prediction accuracy.

Table 4.2. The Smolders, Sanchez and Nisbet dataset are summarised in terms of minimum, maximum and mean value.

	Smolders et al, 1997 (*Lolium perenne*)			Sanchez et al, 1999 (*Agrostis capillaries*)			Nisbet et al, 1999 (*Hordeum distychum*)		
	Min	**Mean**	**Max**	**Min**	**Mean**	**Max**	**Min**	**Mean**	**Max**
OM (%)	3.5	8.4	34	12.6	64.5	96.5	1.2	6	58
Clay (%)	0.5	14.1	36.6	2	12.6	57.6	2.6	16.7	36
exch.Ca ($cmol_c/kg$)	1.4	9.3	35.9	0.5	5.2	34	-	-	-
CEC ($cmol_c/kg$)	5	15.4	37.7	6	12.8	36.7	-	-	-
exch.K ($cmol_c/kg$)	0.15	0.7	1.84	0.1	0.4	0.8	0.09	0.67	2.37
exch.Mg ($cmol_c/kg$)	0.49	1.7	4.32	0.4	2.1	6.4	-	-	-
m_{NH4} (M)	0	$3.7*10^{-4}$	$6.4*10^{-3}$	0	$1.2*10^{-3}$	$4.7*10^{-3}$	-	-	-
pH	4.62	5.3	6.95	3	3.8	6.6	5.2	6.57	8.2
RIP (mmol/kg)	54	1985	5861	5	762.9	6545	-	-	-
TF	37	45	53	0.06	11.5	44.08	$3.1*10^{-3}$	$4.8*10^{-2}$	5.54E-01
m_k (mol/l))	$7.1*10^{-5}$	$2.5*10^{-3}$	$1.2*10^{-2}$	$3.0*10^{-5}$	$2.3*10^{-4}$	$1.6*10^{-3}$	-	-	-

4.2 DESCRIPTION OF THE REVISED CF AND RIP SUB-MODELS

The results of the uncertainty and sensitivity analyses, described in chapter 3, demonstrate that there are differences between the conceptual and the constructed models. Consequently, the CF–m_k and RIP_{clay} models have been revisited with the objective of reducing such discrepancies.

4.2.1 CF–m_k model

As described earlier (4.1.1.1), Absalom assumes a linear correlation between CF and m_k (Figure 4.3a). The variation in the data can also be described by a logistic relationship (equation 4.13; Figure 4.3b), whose performance is statistically similar to the linear model (accounting for 87% and 88% of the data variance respectively).

Nevertheless, the implementation of the two models differs significantly for K^+ concentration in soil solution lower than 0.1 mmol/l and greater than 1mmol/l (Figure 4.3a and 4.3b), as in these two regions the logistic model assumes a lower and a upper limit for the CF-m_k relationship. This appears to resemble more accurately the observed trend (Figure 4.3b) than the linear model (Figure 4.3a).

$$Log(CF) = a + \frac{c}{1 + e^{b*Log(m_k)+m}} \quad (4.13)$$

where:

a, b, c and m are fitted parameters.

m_k is the concentration of K^+ in soil solution (mmol/l).

$$Log(CF) = +k_1 - k_2 Log\left(\min\left(m_k, k_{\lim}\right)\right) \quad (4.14)$$

where:

k_1 and k_2 are fitted parameters.

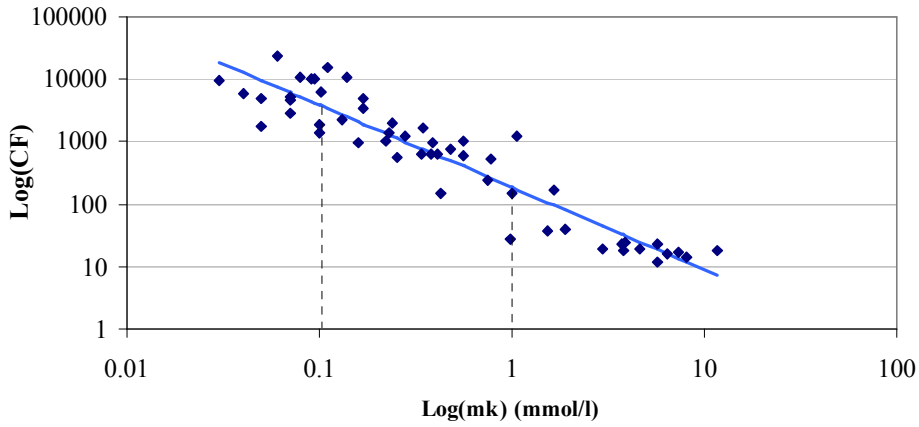
m_k is the concentration of K^+ in soil solution (mmol/l)

k_{lim} is the m_k upper limit (2.4 mmol/l).

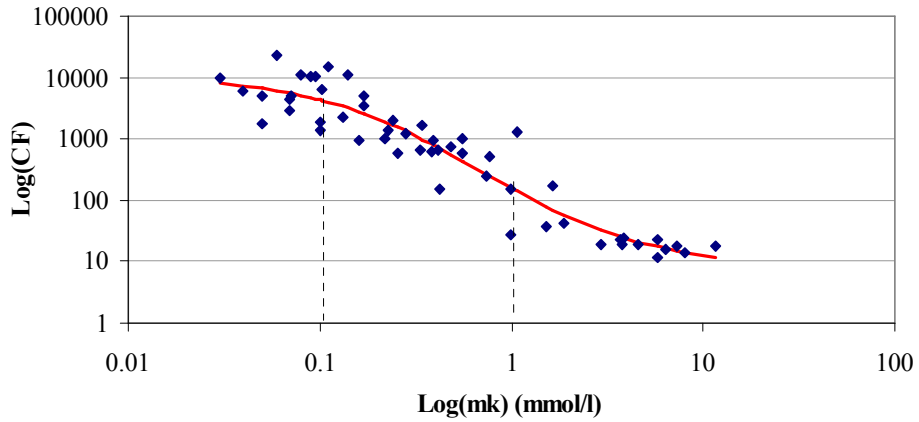
The CF–m_k linear sub-model implemented in the Absalom model (Absalom et al 2001) is based on the sub-model used by Absalom et al (1999) (equation 4.14). In the latter, it is assumed that lowering the K^+ in soil solution strongly increases the Cs uptake by plants. However the negative correlation is established only for values of m_k < 2.4 mmol/l (k_{lim}). Any concentration of potassium higher than 2.4 mmol/l would not affect the plant uptake of Cs, resulting in a constant CF (Smolder et al, 1997). Nevertheless, the use of a factor which limits the CF–m_k correlation has not been implemented in the more recent version (Absalom et al, 2001) as it did not affect the model optimisation with the available data.

The use of the logistic sub-model accounts for a progressive reduction in CF for m_k comprised between 0.1 and 1 mmol/l, and it introduces a plateau for potassium concentration lower than 0.1 mmol/l and greater than 1 mmol/l (Figure 4.3b), where changes in m_k do not affect the CF.

The presence of a plateau, for m_k >1 mmol/l, has been observed in several studies, a comprehensive overview is presented in Zhu et al (2000). Therefore the logistic sub-model is more mechanistically plausible.



(a)



(b)

Figure 4.3 The linear (a) and logistic (b) approach, used to describe the CF–m_k relationship, are compared. The main difference between the two models is that the linear model assumes a negative correlation, between CF and m_k) for any m_k value, while the logistic introduces a maximum and minimum uptake rate for m_k>1 mmol/l and m_k>0.1 mmol/l.

Nonetheless, it is unclear why the Cs uptake rate at high K^+ concentrations is almost unaffected by increasing m_k. Zhu et al (2000) present two hypotheses that could explain this phenomenon. Firstly, at high m_k radiocaesium may be absorbed by plants through different "channels" than those for potassium. Secondly, increasing K^+ concentration in soil solution may result in a remobilisation of radiocaesium through ion exchange reactions, therefore balancing the effects of the negative correlation between CF and m_k (Zhu et al, 2000).

The suggestion that for potassium concentration lower than 0.1 mmol/l, CF is not correlated to m_k is not strongly supported by empirical data. This is due to

experimental limitation, as it is difficult to maintain constant conditions of low potassium concentration in soil solution for the entire experimental period (Zhu et al, 2000). However, it is possible to hypothesize that there is a maximum rate of radiocaesium uptake.

The implementation of a logistic relationship has the benefit of providing a more reliable prediction for soils having high potassium concentration in soil solution, however the drawback is the increase of model complexity. The issue of increasing model complexity is not further discussed here as it is extensively described in Chapter 6.

4.2.2 RIP_{clay} sub-model

The original model (Absalom et al, 2001) implemented the RIP_{clay} sub-model assuming a negligible contribution of organic matter to Cs adsorption. The estimated RIP_{clay} is inversely proportional to θ_{humus} and positively correlated to θ_{clay} (equation 4.10).

Although the hypothesis that clay is the main radiocaesium adsorbent has been proved by several studies (e.g. Facchinelli et al, 2001) a more in-depth analysis of the Smolders and Sanchez datasets, performed as part of this work, showed that the correlation between $Log(RIP_{clay})$ and θ_{clay}–θ_{humus} ratio is weak (Pearson Correlation Coefficient (r) = -0.54).

However, the RIP of the whole soil, i.e. Log(RIP), is significantly correlated to θ_{clay} and θ_{humus}, r is respectively 0.52 and -0.73. Consequently, a multiple-regression analysis was performed and a linear model, which estimates the RIP as function of θ_{clay} and θ_{humus}, has been developed and implemented, equation 4.15.

$$Log(RIP) = -k_{10} + (k_8 * \theta_{clay}) - (k_9 * \theta_{humus}) \qquad (4.15)$$

where

k_8, k_9 and k_{10} are fitted parameters

θ_{clay} and θ_{humus} are respectively the clay and the organic matter soil contents (g/g)

The revised RIP model accounts for 65% of the data variance, while the original sub-model, equation 4.10, only 39% (Figure 4.4).

The increase in model prediction power is due to a combination of factors. Firstly, the revised model can account more accurately for the combined effect of clay and organic matter content on the soil RIP. Secondly, the revisited model uses three optimised global (OG) parameters, while the RIP_{clay} model uses two. The increase in model complexity may enable the model to account more accurately for the data variance.

A more detailed discussion on the relationship between model complexity and goodness-of-fit is given in Chapter 6, therefore it is not discussed further in this section.

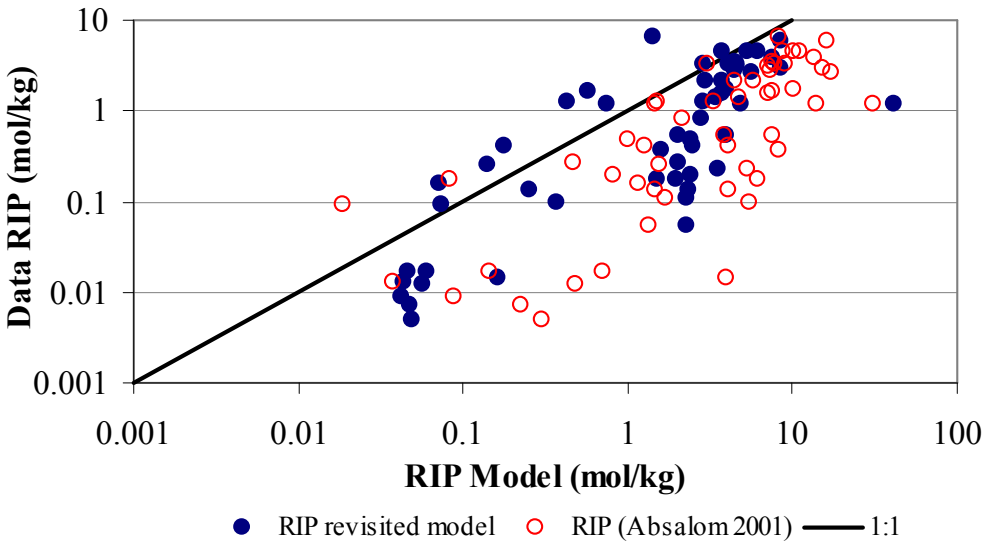


Figure 4.4. The two RIP sub-models, implemented in the Absalom model (○) (Absalom et al, 2001) and in the revised model (●), are compared versus Smolders-Sanchez observations. The solid line represents the 1:1 relationship.

4.3 METHODS

4.3.1 Uncertainty and Sensitivity analysis

An uncertainty and sensitivity analysis has been performed on the revised model output, i.e. TF_{min}, TF_{med} and TF_{max}, using the same methodology introduced and discussed in chapter 3.

4.3.2 Model optimization

The model presented in this work has been optimized applying an extended version of the traditional Levenberg-Marquardt (LM) algorithm. The search for the function minimum is performed using the LM approach, however it incorporates the basic feature of a Simulated Annealing method, i.e. the "uphill" sampling, which enables the search to avoid local minimum, increasing the confidence that the result refers to the global minimum. A detailed overview of the LM algorithm can be found in Fletcher (1987), while a comprehensive introduction to the Simulated Annealing method can be found in Gershenfeld (1999).

This method has been preferred to other methodologies proposed in the literature (Gershenfeld, 1999 and Mitchell, 1995 for an overview) as it represents a good compromise between reliability in searching for global minimum and computational cost.

The function that has been minimised is the Residual Sum of Square (RSS), equation 4.16, which is a measurement of the degree of agreement between prediction (x) and empirical data (y): the lower is the RSS, the higher is the model accuracy.

The optimisation was performed simultaneously on three model variables, log(mk), log(kd) and log(TF) (Table 4.3). This optimisation strategy can be defined as Multivariable Optimisation (MVO) as more than one variable is considered in the optimisation procedure. The use of the MVO ensures that key model components provide realistic estimations, which can increase the model applicability. However, the MVO method could be criticized since the covariance between model variables is not taken into account. Nonetheless, this approach has been successfully used in

several works (e.g. Absalom et al, 1999 and Absalom et al, 2000). Therefore, its use has been regarded as consistent with the previous publications on the subject.

$$RSS = \sum_{i=1}^{n} \left(x_i - y_i \right)^2 \tag{4.16}$$

where

x_i and y_i are respectively the model prediction and the empirical data

n is the total sample number (over m_k, k_d and TF).

4.4 RESULTS AND DISCUSSION

4.4.1 Model accuracy tests

The accuracy of the model predictions has been tested comparing the estimated TF with empirical data of the Smolders and Sanchez studies and NRPB (Nisbet et al, 1999) datasets. However, the assessment of the prediction accuracy using the Smolders-Sanchez and NRPB dataset cannot be considered a "blind validation". The former has been used for the model optimisation (Table 4.4) and the latter refers to measurements of barley (*Hordeum distychum*) contamination (Bq/kg). Consequently, it might be beyond the model capabilities to provide an accurate prediction of Cs activity concentration in barley using OG parameters optimised for pasture. Therefore, in order to test and compare the degree of accuracy of the models considered, i.e. Absalom and the revised model, the CF–m_k relationships, equation 4.3 and 4.13 respectively, have been refitted to the NRPB data, using an approach similar to the one of Absalom et al (1999) (Table 4.5). This optimisation has been performed only on the CF–m_k sub-model as it is the only model compartment which is plant species specific.

Table 4.3 The Residual Sum of Square (RSS) of the original (Absalom et al, 2000) and the revised model are compared for each model variable considered in the optimisation procedure.

	Absalom 2000			Revised model		
	log(TF)	log(K_d)	log(m_k)	log(TF)	log(K_d)	log(m_k)
Smolders dataset	5.35	4.03	3.71	5.24	8.13	3.40
Sanchez dataset	11.31	10.44	4.30	5.61	7.86	4.64
Total	16.66	14.47	8.01	10.86	15.99	8.04

Table 4.4 Values of fitted and constants (*) model parameters used in the revised model and ± the parameter standard error.

Model parameter	Value	Equations
a	1.49 ± 0.16	3.13
b	16.79 ± 12.11	3.13
c	1.94 ± 0.20	3.13
m	3.18 ± 0.078	3.13
k_G^{clay}	3.79 ± 1.08	3.7
k_G^{humus}	1.87 ± 0.42	3.7
k_3	3.368 *	3.5
k_4	0.16 *	3.5
k_5	-34.66 *	3.6
k_6	29.72 *	3.6
k_7	0.24 ± 0.13	3.9
k_8	2.52 ± 0.53	3.14
k_9	1.92 ± 0.24	3.14
k_{10}	0.40 ± 0.13	3.14
CEC_{clay}	50 *	3.4
P_{fast}	0.814 *	3.12
k_{fast}	0.0019 *	3.12
k_{slow}	0.00019 *	3.12

This approach is in contrast to that of Absalom et al (2001), who estimated the accuracy of model prediction using the NRPB barley data without re-optimisation. It was assumed that the soil-to-plant TF for grass and barley are comparable. Nevertheless, this is a controversial assumption, as measurements of TF for barley and grass have been reported to differ by one order of magnitude (IAEA, 1994). However, TF is a function of several environmental factors, for instance pH, clay, organic matter, and the grass-barley difference could be explained by the TF variability, which is known to be high (IAEA, 1994; Frissel et al, 2002). Nonetheless, in the present work TF for grass and barley are not assumed to be comparable as is suggested in the IAEA handbook.

Table 4.5 OG parameters of the Absalom and the revised model CF–m_k relationship refitted on the NRPB dataset.

	Model parameter	Value
Revisited model	*a*	0.13 ± 0.16
	b	1.29 ± 0.91
	c	4.57 ± 3.20
	m	2.89 ± 0.13
Absalom model	k_1	2.03 ± 0.38
	k_2	1.58 ± 0.12

Table 4.6 R^2 estimated for the revised model and the Absalom model prediction versus Smolders-Sanchez dataset.

	Absalom 2000	Revised model
log(m_k)	0.70	0.70
log(k_d)	0.74	0.72
log(TF)	0.79	0.86

4.4.1.1 Model validation: Smolders-Sanchez dataset

The predicted m_k, k_d and TF of the Absalom model (Absalom et al, 2001) and the revised version, proposed in this work, have been compared with measurements from the Smolders-Sanchez dataset, Figure 4.5. There is not a substantial difference between the two model predictions for log(m_k) and log(k_d), Table 4.6. Nevertheless, a better fit is evident for the predicted TF, r^2 is 0.79 and 0.86 for the original and revised versions, respectively.

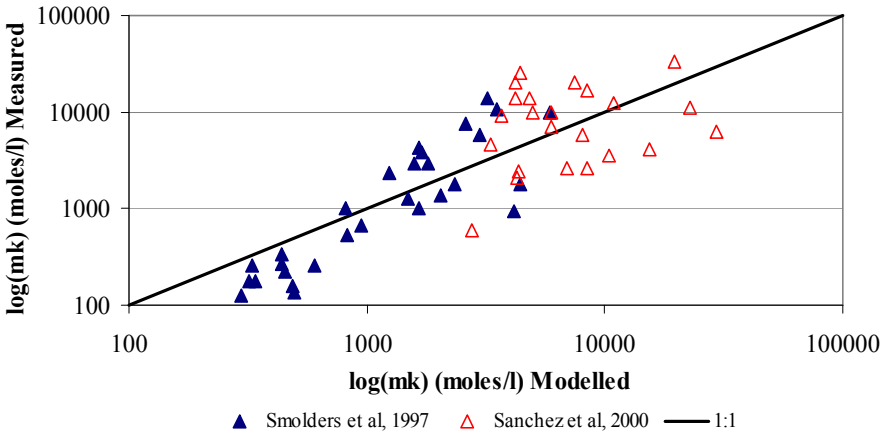
Figure 4.5e and 4.5f illustrate that the improvement in model prediction capability (Table 4.5) is mostly attributed to organic rich soils, which are part of the Sanchez dataset. The estimated RSS for the predicted TF (Table 4.3) is 16.66 and 10.86 for the original and revised model respectively.

4.4.1.2 Model validation: NRPB datasets.

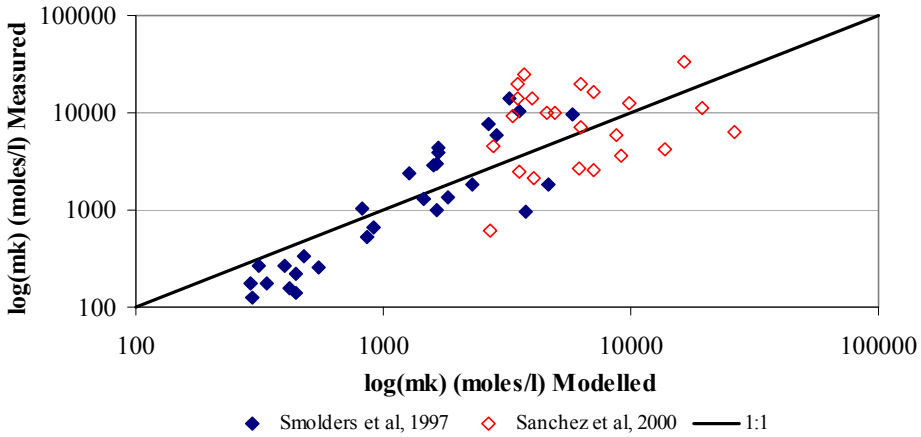
The results of the validation tests, for the Absalom model and the revised version, performed with the NRPB dataset are presented in figure 4.6.

As outlined earlier the two models have been optimised with the Smolders-Sanchez dataset (grass data). In order to compare the NRPB data and the model predictions the Absalom and revised models CF-m_k relationships have been re-optimised with the barley data. The revised model predictions show a moderate increase in the degree of agreement with the observations, figure 4.6a and 4.6b. The new model is characterised by a coefficient of correlation (r^2) equal to 0.55 while the r^2 for the Absalom model is 0.49.

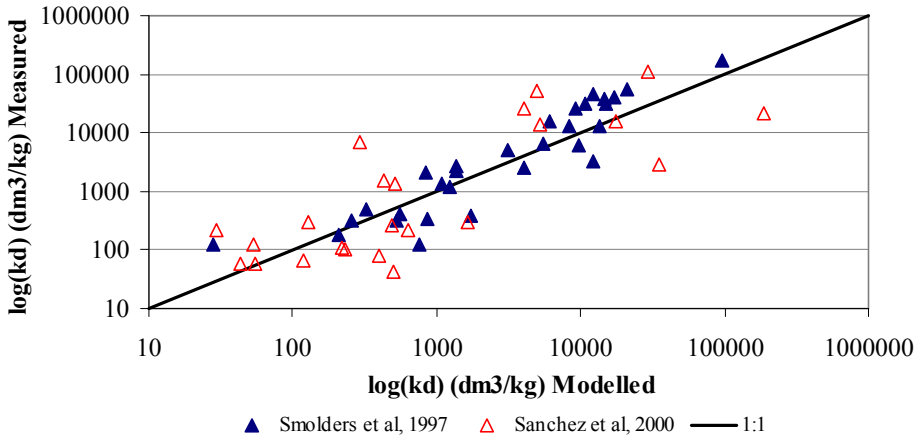
However, the correlation analysis between the residuals and the soil characteristics did not show any significant trend that could explain the results obtained, Figure 4.6. Therefore, the increase of the model accuracy could be simply due to the use of the logistic function for the CF–m_k relationship, which is formed by four parameters, which consequently can account more efficiently for the empirical data variation than the linear model used in Absalom et al (2001).


log(mk) (moles/l) Measured
100000
10000
1000
100
100
1000
10000
100000
log(mk) (moles/l) Modelled
Smolders et al, 1997
Sanchez et al, 2000
1:1

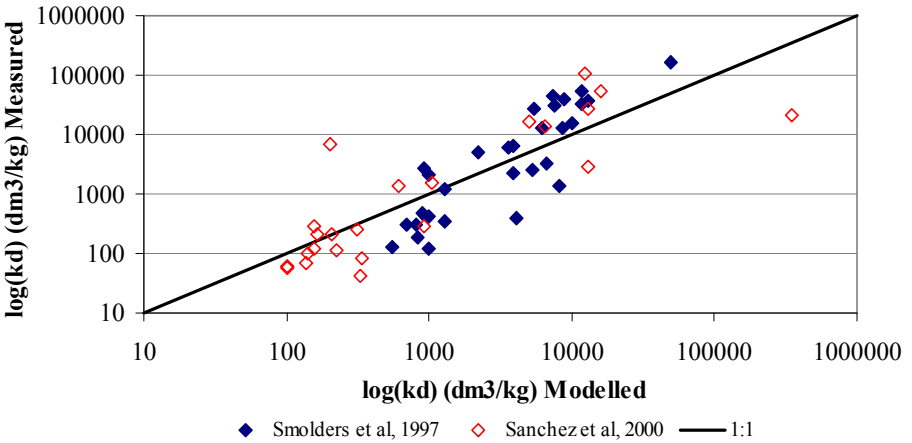
(a)


log(mk) (moles/l) Measured
100000
10000
1000
100
100
1000
10000
100000
log(mk) (moles/l) Modelled
Smolders et al, 1997
Sanchez et al, 2000
1:1

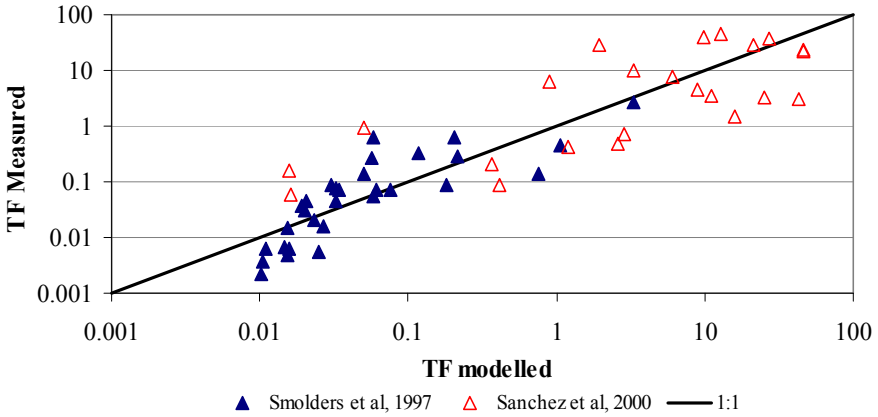
(d)


log(kd) (dm3/kg) Measured
1000000
100000
10000
1000
100
10
10
100
1000
10000
100000
1000000
log(kd) (dm3/kg) Modelled
Smolders et al, 1997
Sanchez et al, 2000
1:1

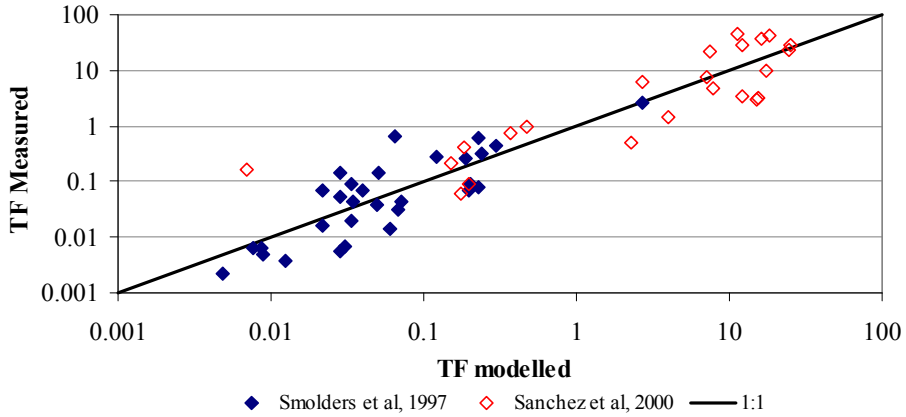
(b)


log(kd) (dm3/kg) Measured
1000000
100000
10000
1000
100
10
10
100
1000
10000
100000
1000000
log(kd) (dm3/kg) Modelled
Smolders et al, 1997
Sanchez et al, 2000
1:1

(e)

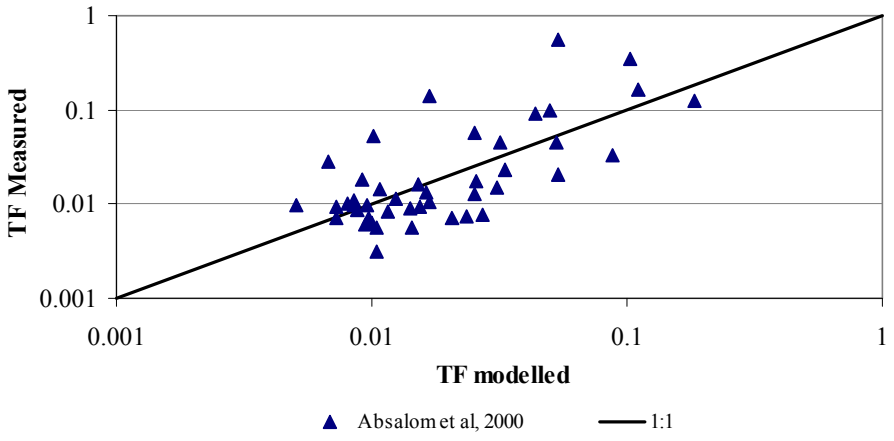


(c)

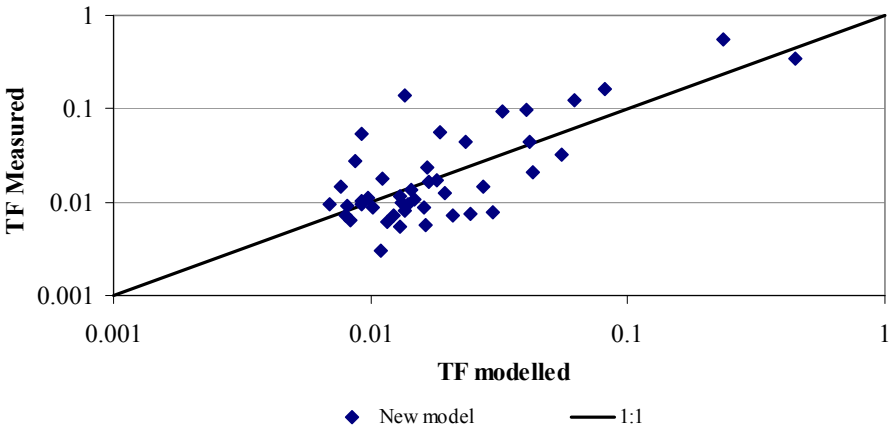


(f)

Figure 4. 5 Comparison of the revised (a-c; ▲) and the Absalom model (d-f; ♦) predictions for log(mk), log(kd) and log(TF), using Smolders et al (1997) (filled symbols) and Sanchez et al (1999) (open symbols).



(a)



(b)

Figure 4.6 The Absalom model (▲) and the revisited model (♦) predictions of barley contamination are compared against the NRPB dataset.

4.4.2 Uncertainty and Sensitivity analysis

The revised model has been tested for uncertainty and sensitivity considering the effect of input and model parameters on the predicted soil-to-plant TF_{min}, TF_{med} and TF_{max}.

The predicted TF are lognormal distributed, figure 4.7. The NR_{75} values show that the revised model has a 50% reduction in the prediction uncertainty (Table 4.7), which is due to the more even model sensitivity to model factors, and in particular to the significant influence of input data (Figure 4.8).

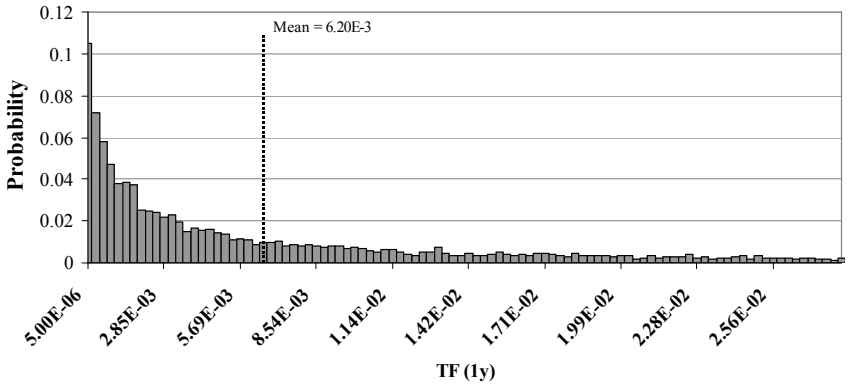
Table 4.7 The NR_{75} values for the three TF categories are reported for 1, 10 and 20 years.

	Revised Model			Absalom et al 2001		
	1 year	10 years	20 years	1 year	10 years	20 years
TF_{min}	2.08	2.11	2.06	4.59	4.58	4.67
TF_{med}	2.38	2.39	2.38	6.12	6.01	6.12
TF_{max}	2.61	2.74	2.85	6.38	6.84	6.89

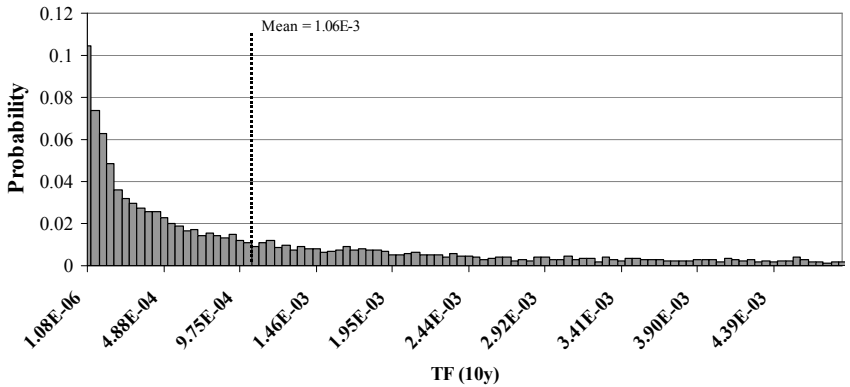
The prediction of TF_{min} is strongly dominated by the CF–m_k and RIP sub-model. The parameters *a* and *b* account for 40% of the prediction variance (Figure 4.8b). This effect is mostly due to the use of a logistic model for the CF–m_k relationship, as it assumes a poorer correlation between CF and m_k for value of m_k > 1 mmol/l (Figure 4.3b). Consequently, m_k uncertainty does not affect significantly the CF uncertainty, which is only affected by the OG parameters, i.e. a, b, c and m part of the logistic sub-model.

The sensitivity of the model to clay, affecting 14% of the uncertainty, at low TF is consistent with the assumption that the Cs availability is reduced by clay, since it specifically adsorbs the radiocaesium, which is therefore not available for plant uptake. This is further emphasised by the high model sensitivity to the RIP model parameters, i.e. k_8 and k_{10}, which account for 35% of the prediction uncertainty.

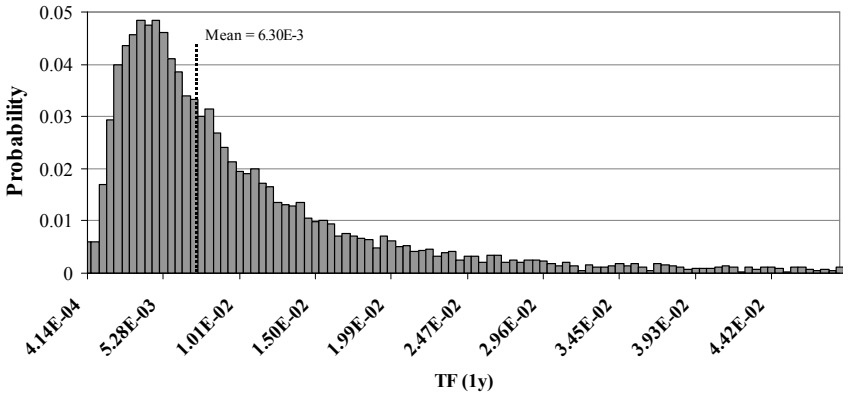
At a medium value of TF (TF_{med}), the model uncertainty is dominated by the clay content, CV equal to 20%. The organic matter only partially influences the model prediction, CV equal to 4%. This is in accord with the assumption that OM has a significant role in Cs availability only in organic rich soils, where the effect of OM can overcome the clay influence. Nevertheless, the CF–m_k and RIP sub-models still have a key role in the model prediction uncertainty.


Mean = 6.20E-3
Probability
0.12
0.1
0.08
0.06
0.04
0.02
0
5.00E-06
2.85E-03
5.69E-03
8.54E-03
1.14E-02
1.42E-02
1.71E-02
1.99E-02
2.28E-02
2.56E-02
TF (1y)

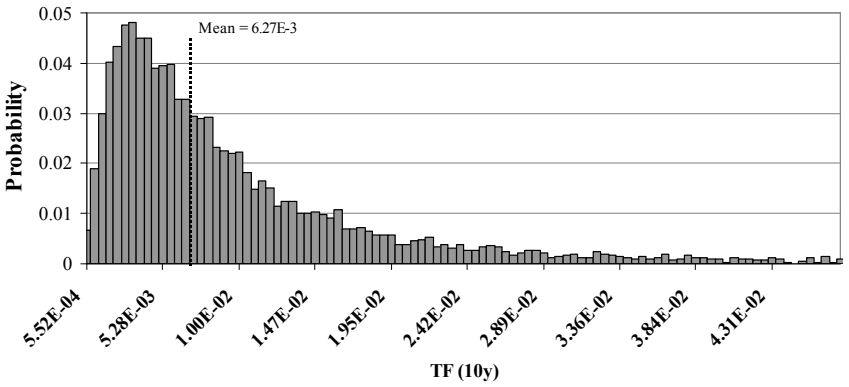
(a)


Mean = 1.06E-3
Probability
0.12
0.1
0.08
0.06
0.04
0.02
0
1.08E-06
4.88E-04
9.75E-04
1.46E-03
1.95E-03
2.44E-03
2.92E-03
3.41E-03
3.90E-03
4.39E-03
TF (10y)

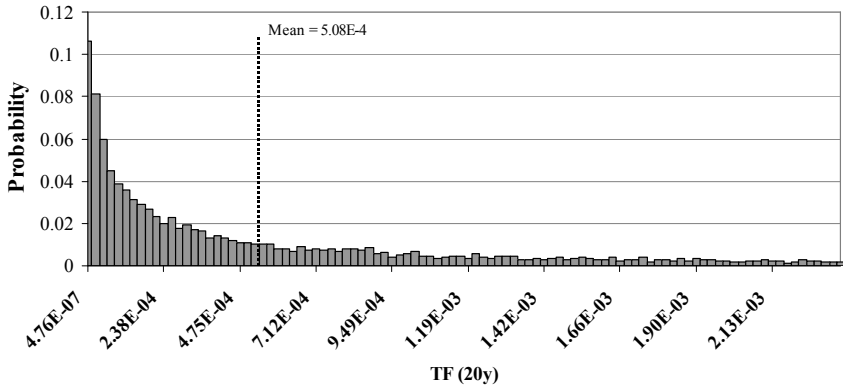
(b)


Mean = 6.30E-3
Probability
0.05
0.04
0.03
0.02
0.01
0
4.14E-04
5.28E-03
1.01E-02
1.50E-02
1.99E-02
2.47E-02
2.96E-02
3.45E-02
3.93E-02
4.42E-02
TF (1y)

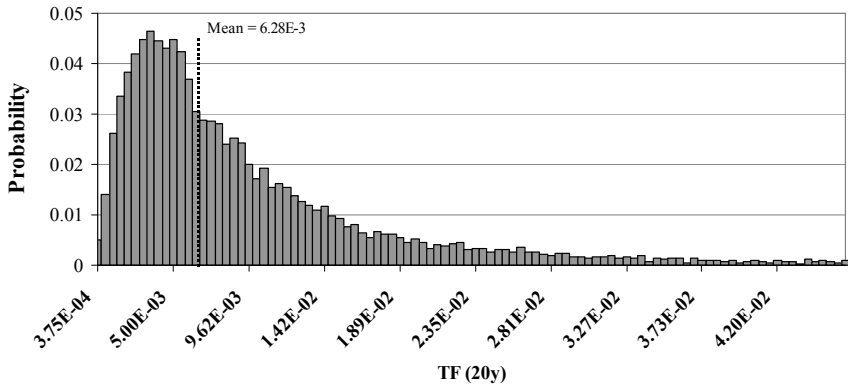
(d)


Mean = 6.27E-3
Probability
0.05
0.04
0.03
0.02
0.01
0
5.52E-04
5.28E-03
1.00E-02
1.47E-02
1.95E-02
2.42E-02
2.89E-02
3.36E-02
3.84E-02
4.31E-02
TF (10y)

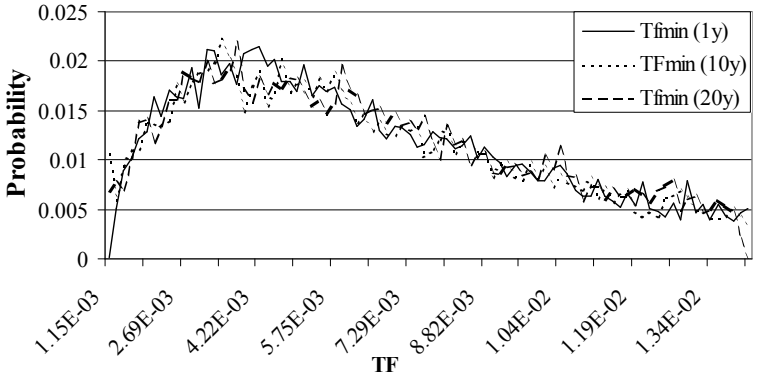
(e)



(c)



(f)



(g)

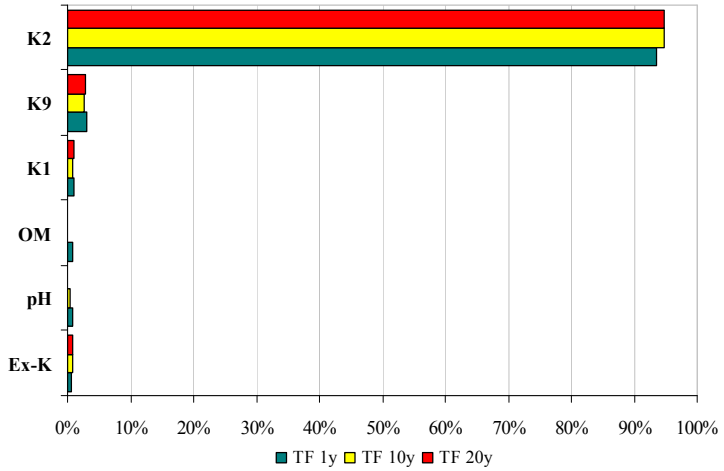
Figure 4.7 Probability density function of TF_{min} for 1, 10 and 20 years after contamination for the Absalom (a-c) and revised model (d-f). In addition the probability density function of TF_{min} are compared (g).

It is interesting to observe that the k_7 parameter, which is the m_{NH4} coefficient in the k_d^{clay} relationship (equation 4.9), accounts for 7% of the prediction uncertainty, while in the TF_{min} results its effect was negligible. This is due to the increase in the sensitivity of the model to the NH_4^+ in soil solution, CV equal to 1.2%. The effect of ammonium can be generally considered negligible, however in the scenario considered, Sanchez dataset (Sanchez et al, 1999), the NH_4^+ concentration in soil solution is rather high, $7.8*10^4$ $cmol_{eq}/kg$, therefore it is reasonable that the model shows a sensitivity to m_{NH4}.

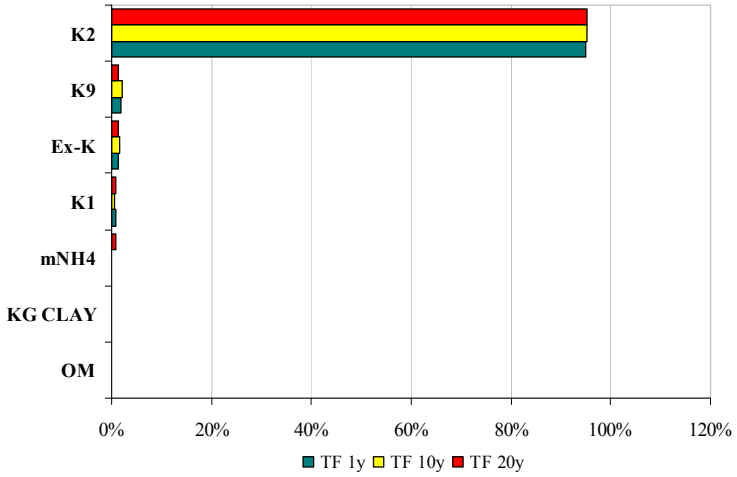
The TF_{max} sensitivity analysis shows that the model is highly sensitive to the organic matter content, CV equal to 40%, and to the k_9 parameter, CV equal to 20%. In organic rich soils the effect of organic matter in Cs availability to plant uptake has been observed to be more significant than the clay content, and the model accounts for this. The scenario considered is based on 95% organic matter and 4.5% clay (Table 4.2).

Time does not appear to be an influential parameter on the prediction uncertainty for the three TF categories as the NR_{75} at 1, 10 and 20 years after contamination can be regarded as constant, Table 4.7.

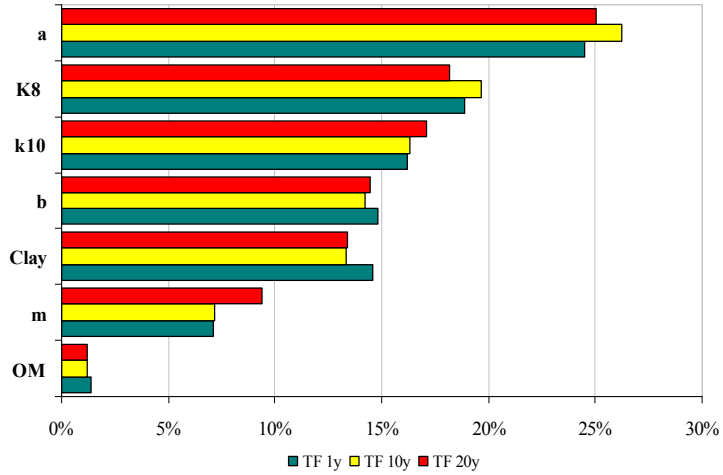
The revised model is characterised by a lower level of uncertainty than the Absalom model (Absalom et al 2001): the NR_{75} are generally reduced by 50%, and this is mainly due to the more balanced model sensitivity. In addition the sensitivity analysis has demonstrated that the revisited model developed has a lower sensitivity to optimised global parameters while the sensitivity to local parameters has increased. In addition, the model sensitivity is consistent with the general understanding of processes which affect the Cs availability to plant uptake. Therefore, the revisited model shows a higher level of resemblance with the conceptual model than the Absalom model.



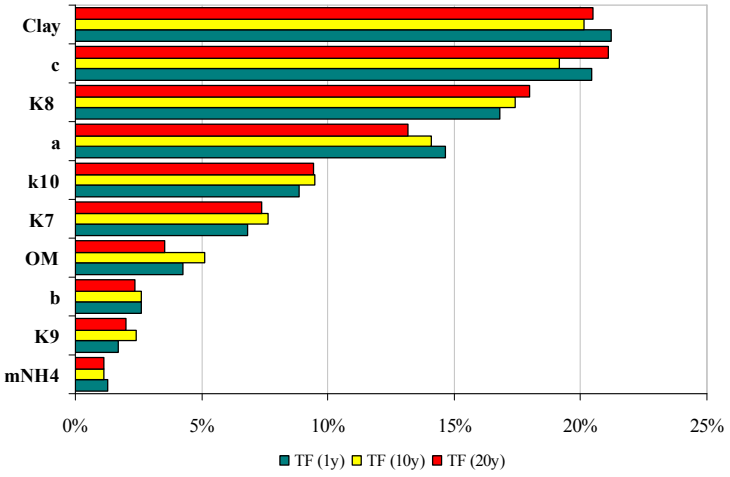
(a) TF_{min}



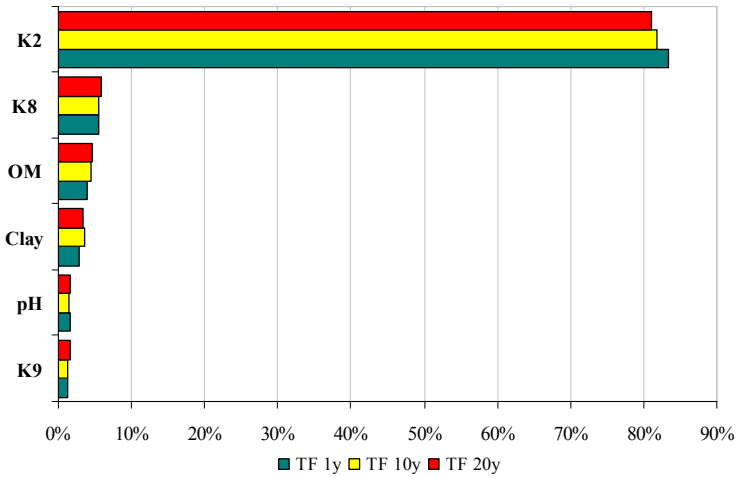
(b) TF_{med}



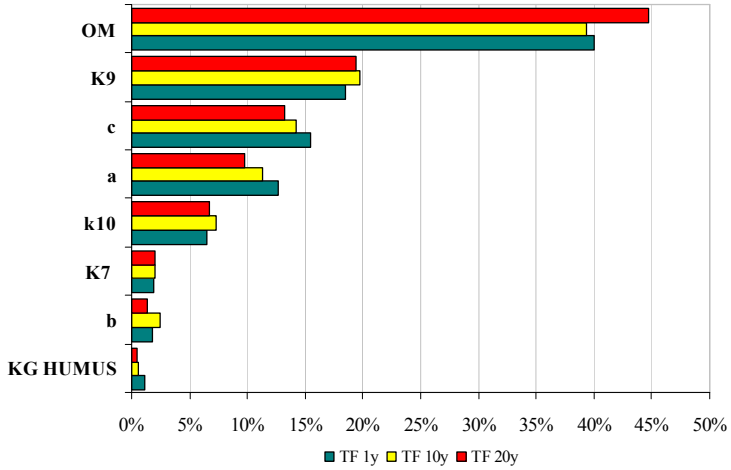
(d) TF_{min}



(e) TF_{med}



(c) TF_{max}



(f) TF_{max}

Figure 4.8 Sensitivity analysis results (CV, %) of the Absalom (a-c) and the revised model (d-f)

4.4 CONCLUSIONS

The aim of this work was to perform a re-evaluation of the model developed by Absalom et al (2001) based on the sensitivity analysis results, in order to reduce the discrepancies between the implemented and conceptual model.

A revised CF and RIP relationship has been implemented and the model has been re-parameterised.

The resulting model has been tested for accuracy; the predicted TF has been compared with the combined Smolders-Sanchez dataset and with measurements of barley (Bq/kg), from the NRPB database (Nisbet, et al, 1999).

While the proposed model does not show any significant difference with the Absalom model in terms of prediction accuracy, it has a lower level of uncertainty than the later (the NR_{75} shows a 50% reduction). The model sensitivity is consistent with the understanding of the behaviour of Cs in soil, which therefore increases the resemblance between the revised and conceptual models.

4.5 SUMMARY

Chapter aim: Revaluate the Absalom model (Absalom et al, 2001) in order to increase the similarities between the conceptual and the constructed model.

- A revised CF and RIP relationship, based on a detailed analysis of the Smolders-Sanchez data, has been implemented in the model.
- Four conclusions can be established based on the comparison between the revised and the Absalom models:
    a) The revisited model does not show higher prediction accuracy than the Absalom model.
    b) The level of uncertainty is reduced by circa 50%.
    c) The sensitivity analysis shows that local parameters play a key role in the model uncertainty.
    d) Finally, there is a significant agreement between the model sensitivity results and the general understanding of processes

affecting Cs soil-to-plant transfer. Therefore, the revised model is characterised by a higher level of resemblance with the conceptual model than the Absalom model.

5. MODEL ACCURACY: COMPARISON OF PREDICTIONS TO OBSERVATIONS

Mathematical models are generally used to simulate the behaviour of a system and often to predict the potential effects of an unusual event. If the event considered has caused significant changes in the system, the model predictions can be used to estimate the most efficient approach that should be undertaken in order to restore the system's stability.

In the particular case of this work the systems considered are the European rural and semi-natural environments and the unusual event is radiocaesium deposition following the Chernobyl accident. The models considered can provide important information on the behaviour of ^{137}Cs within the food chain, which can be used to plan a countermeasure strategy to reduce doses to humans.

Models are developed using the current understanding of natural systems, which is far from complete. Thus, as highlighted previously in this work, model predictions have to be verified and model prediction errors quantified. Nevertheless it is not possible, for most models, to ensure that the constructed model operates as intended in every possible scenario. In other words, rarely can a model be fully verified. More commonly the model prediction power is quantified only for a limited number of scenarios; nevertheless, this provides important guidelines for the application of the model.

The aim of this chapter is to establish the degree of similarity between the predictions of the models considered and independent data. To achieve this, two scenarios have been selected: rural (south Finland) and semi-natural (a state forest in south Germany). Measurements of ^{137}Cs activity concentration in four food-products have been used to test the rural models (TEMAS rural and SAVE rural) while the semi-natural models (RIFE1, FORM and SAVE semi-natural) have been tested using data of ^{137}Cs concentration in roe deer (Bq/kg) from Bad Waldsee state forest, Germany.

The chapter is divided into three sections. An introductory part which provides a description of the two scenarios considered, a second section, where the model-scenario comparison results are presented and the final part where results are critically discussed.

5.1 INTRODUCTION

The south Finland and Bad Waldsee datasets have been chosen for the model-scenario test because they have not been used for any development or calibration of any of the considered models, therefore they can be regarded as independent data. Consequently, this test can be classified as a blind validation.

Table 5.1 Models considered have been tested for prediction accuracy comparing their outputs against measurements from two scenarios south Finland and Bad Waldsee state forest, rural and semi-natural environment respectively.

Models	Ecosystem	Foodstuffs
SAVE rural – Absalom model	Rural	Dairy milk (Bq/l) Beef (Bq/kg)
TEMAS rural	Rural	Pork (Bq/kg) Cereals (Bq/kg)
SAVE semi-natural	Semi-natural	
RIFE1	Semi-natural	Roe deer (Bq/kg)
TEMAS semi-natural – FORM	Semi-natural	

The experimental values used refer to radiocaesium concentrations in food products, due to the radionuclide deposition following the Chernobyl accident in 1986.

Four food-stuffs have been used to test the rural models: dairy milk (Bq/l), beef (Bq/kg), pork (Bq/kg) and cereals (Bq/kg). The semi-natural model comparison has been limited to roe deer contamination (Bq/kg). This is not due to lack of data; the Bad Waldsee dataset includes also measurements of ^{137}Cs activity concentration in wild boar (Bq/kg) and mushrooms (Bq/kg). However, these two endpoints are characterised by a high variability, due to the high influence of local soil characteristics and growth stage (Baeza et al 2005; Linkov et al, 2000; Kammerer et al, 1994) for the former and the complex diet for the latter (Vilic et al, 2005). Therefore, these data have not been included in the test, as it was not considered a representative assessment of the potential model accuracy.

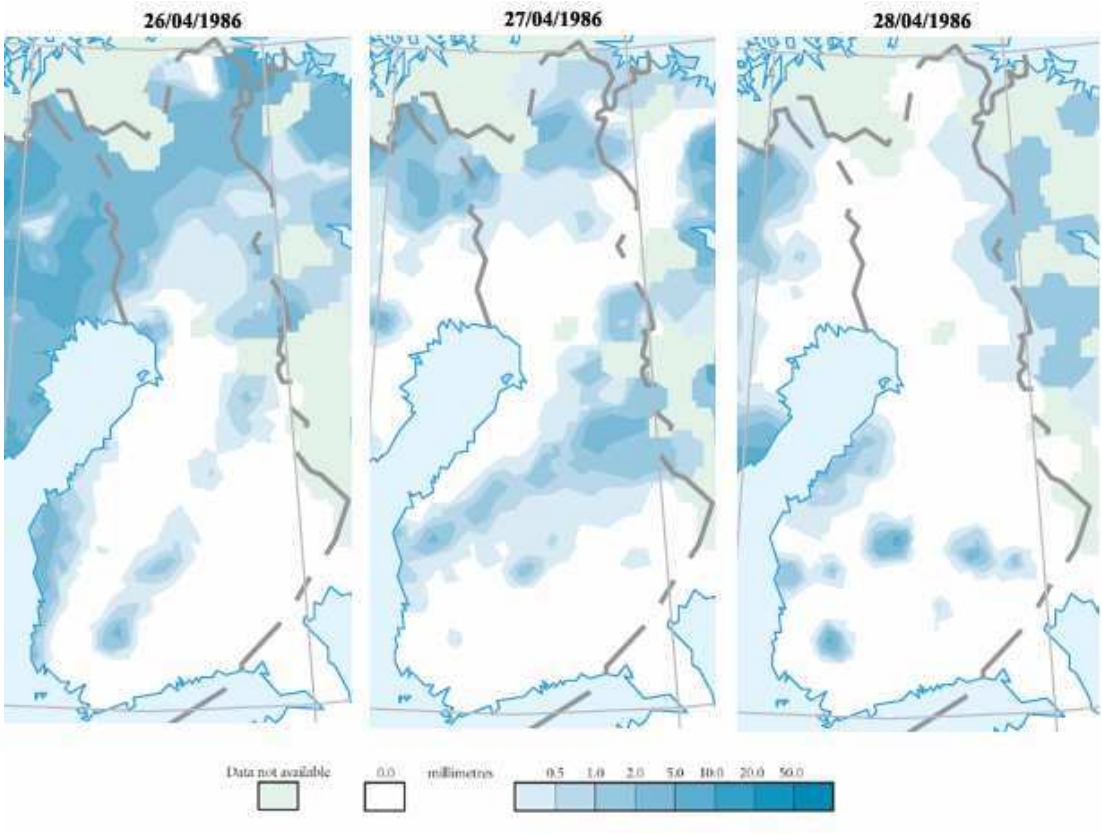
5.1.1 Rural scenario: South Finland

The data from the south Finland region refer to 4 years of measurements (1986 – 1990) of ^{137}Cs activity concentrations in food products, caused by the Chernobyl accident. This dataset has been previously used for an IAEA/CEC co-ordinated research programme on the validation of environmental model predictions (VAMP). Therefore, most of the scenario details, provided in this work, are taken from the VAMP final report (IAEA, 1996).

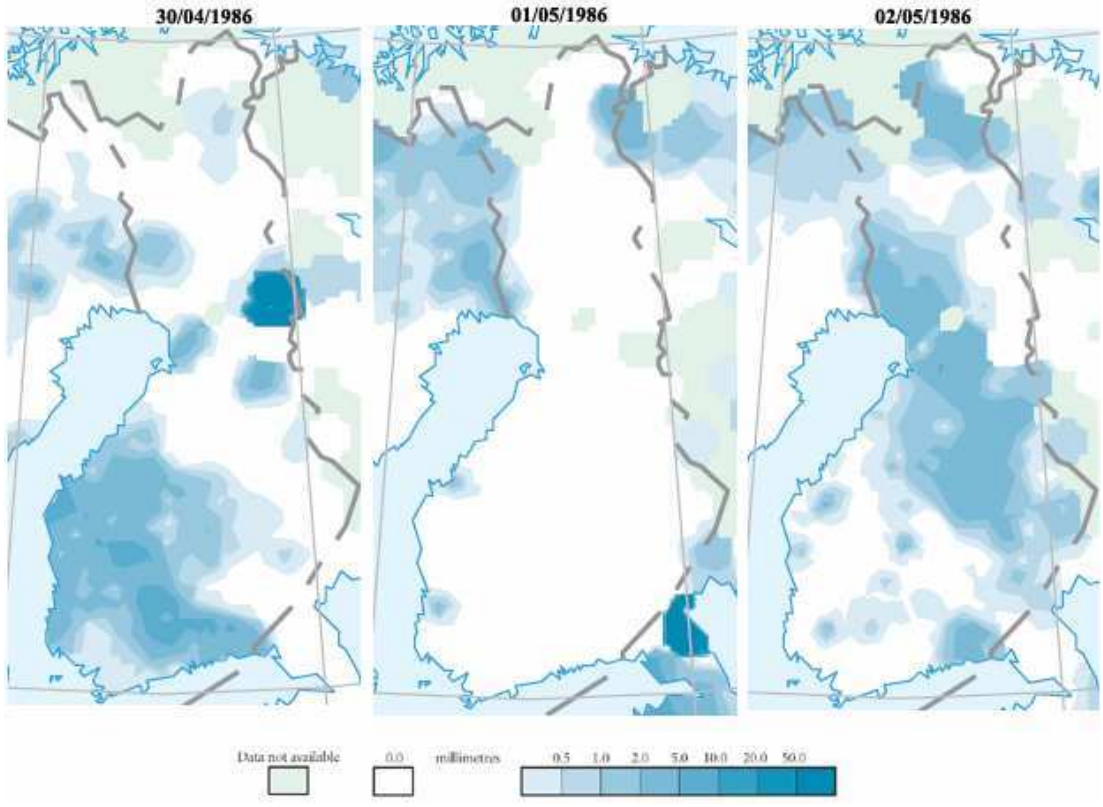
The test area comprises the nine southernmost provinces of Finland, and has a total area of 176000 km^2, with a population of 4.3 million people.

Most radioactive fallout received in Finland after the Chernobyl accident occurred between the 27th and 29th of April 1986 and most of it was deposited with heavy showers, wet deposition (Figure 5.1a and 5.1b), which caused uneven spatial distribution (Figure 5.2). The highest values were recorded in the AGR12, AGR4 and AGR5 areas, 46.6, 26.0 and 43.8 (kBq/m^2) respectively (Figure 5.2, Figure 5.3 and Table 5.2).

The deposition continued with varying intensity, due to climatic conditions, until May 12th. However at the end of April the growing season had not yet started, consequently crop contamination was primarily by soil uptake rather than foliar absorption. Additional details, regarding the radioactive plume, are described in Apsimon et al, 1989.



(a)



(b)

Figure 5.1 Precipitation (mm) on Finland recorded for period from the 26th to the 28th of April and from the 30th of April to the 2nd of May. (De Cort, 2001)

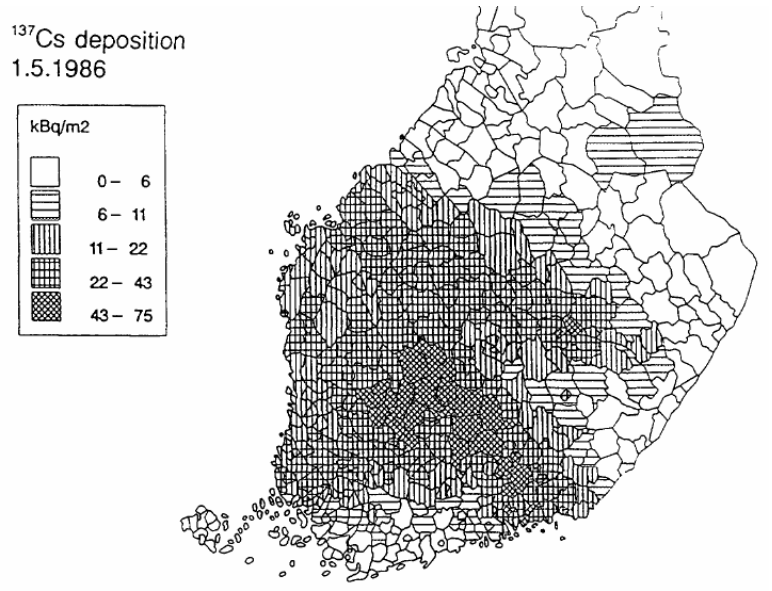
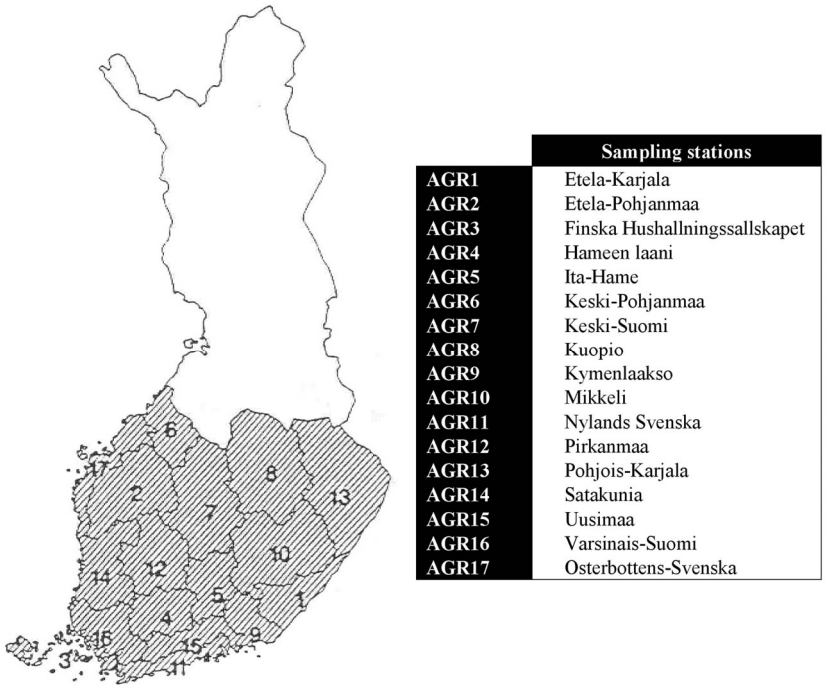


Figure 5.2 ^{137}Cs deposition distribution (kBq/m^2) on south Finland after the Chernobyl accident. ^{137}Cs spatial distribution has been derived using the measurements obtained from the deposition sampling stations and the relative soil vertical distribution (IAEA, 1996).



	Sampling stations
AGR1	Etela-Karjala
AGR2	Etela-Pohjanmaa
AGR3	Finska Hushallningssallskapet
AGR4	Hameen laani
AGR5	Ita-Hame
AGR6	Keski-Pohjanmaa
AGR7	Keski-Suomi
AGR8	Kuopio
AGR9	Kymenlaakso
AGR10	Mikkeli
AGR11	Nylands Svenska
AGR12	Pirkanmaa
AGR13	Pohjois-Karjala
AGR14	Satakunia
AGR15	Uusimaa
AGR16	Varsinais-Suomi
AGR17	Osterbottens-Svenska

Figure 5.3 The seventeen sampling stations. Each station represents the mean contamination for each region (IAEA, 1996).

Table 5.2 Regions characteristics and ^{137}Cs deposition for the considered agricultural areas, estimations refer to May 1986 (IAEA, 1996).

Area	Land (km^2)	^{137}Cs mean value (kBq/m^2)	Cultivation (km^2)	Pasture (km^2)
AGR1	5676.6	6.2	640.27	27.42
AGR2	14512	25.7	2581.07	60.68
AGR3	3075.9	3.8	288.06	56.51
AGR4	6861.2	26	1522.37	60.94
AGR5	5715.8	43.8	668.1	49.89
AGR6	6889.6	14.5	671.56	16.52
AGR7	14883	27.8	975.87	90.79
AGR8	16511.3	11.4	1490.25	116.7
AGR9	5106.7	33.5	852.76	28.43
AGR10	14431	13.9	939.49	89.61
AGR11	4551.6	7.7	724.22	27.71
AGR12	9363.5	46.6	1056.16	77.89
AGR13	17782.3	2.3	1044.98	105.69
AGR14	10502.5	25.6	1745.45	68.66
AGR15	5346.9	18.5	1328.57	45.59
AGR16	8446.9	16.4	2356.63	61.14
AGR17	6976.5	22	1043.86	38.38

5.1.1.1 Ground contamination

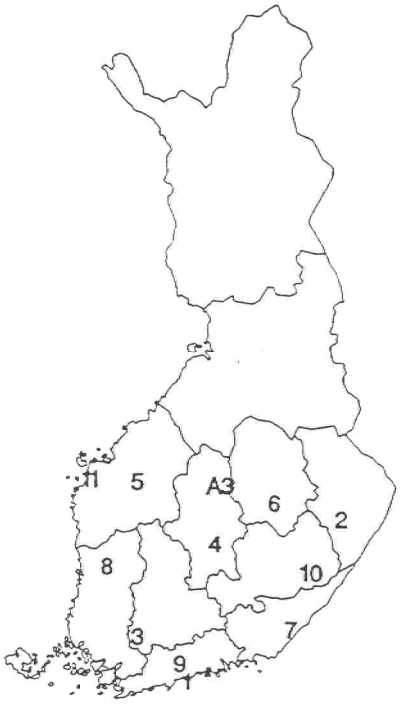
The deposition measurements, taken from eleven sampling stations (Figure 5.4) and the radiocaesium vertical distributions in soil profiles have been used to reconstruct the initial 1986 ^{137}Cs deposition, (Figure 5.2). Urban areas have been excluded as the survey aimed to establish the level of contamination of rural and semi-natural regions, which affects food production.

The locations of the deposition sampling stations are shown on the map in figure 5.4. Four of the sites (DEP1, DEP3, DEP6 and DEP10) are on cultivated grass-covered areas or lawns. Samplers at stations DEP2, DEP5 and DEP9 are surrounded by a flat open area covered with wild surface vegetation. The other stations are mostly located at the edges of open fields, surrounded by large forest areas.

5.1.1.2 General meteorological characteristics and climatic conditions of Finland

The annual duration of the snow cover in open fields is generally between four and six months. However, the mean snow period is from the 1st of November to 30th of April, having the maximum snow cover, expressed in terms of water equivalent, of 100-150 mm between the 15th of March and the 1st of April.

Ground frost usually stops between the middle of April and the middle of May. The average date of ice-break-up in lakes varies between 30th April and 10th May and that of freeze-up between November 15th-30th. In the spring of 1986 most lakes were clear of ice before April 27th.



	Sampling stations
DEP1	Helsinki
DEP2	Joensuu
DEP3	Jokioinen
DEP4	Jyvaskyla
DEP5	Kauhava
DEP6	Kuopio
DEP7	Lappeenranta
DEP8	Niinisalo
DEP9	Nurmijarvi
DEP10	Savonlinna
DEP11	Vaasa
AIR1 (1)	Helsinki
AIR2 (9)	Nurmijarvi
AIR3 (A3)	Viitasaari

Figure 5.4 Deposition and air sampling stations. The outlined areas do not correspond to the country regions (IAEA, 1996).

The sum of effective temperature (the sum of daily average temperature exceeding 5 °C) varies in the test region between 900 and 1350 °C estimated from 30 years of measurements. The mean temperature of the growing season is 12-13 °C and the length of growing season is 160-180 days.

Precipitation is received rather evenly during the whole year. However, rain deficiency may occur at the beginning of the growing season. The annual precipitation in the regions considered generally varies between 450-750 mm. During periods of low rainfall, irrigation is used, especially in sub-areas AGR16 (rainfall deficiently of 80-100 mm in May – July) and in AGR3, AGR11 and AGR14 (rainfall

deficiency of 60-80 mm) (Figure 5.3). Irrigation of fields is most common for vegetable production, but it is used also for cultivation of cereals. Local surface water systems are used as sources of irrigation water.

The best climatic conditions for growing plants are in the south-western and southernmost coastal areas which belong to the hemi-boreal zone.

5.1.1.3 Soil

The most common soils in Finland are *Orthic Podzols*, which are present in 49% of the territory and the *Dystric* and *Eutric Histosols* which cover 28% of the country, (Sippola and Yli-Halla, 1998). The former is common in the circumpolar belt approximately from the artic circle southwards to the latitude of St Petersburg (Europe) and to the northern shores of the Great Lakes in North America (Sippola and Yli-Halla, 1998), while it is rare in the southern continent (Bridges, 1997). On the other hand, the latter is a peat organic rich soil (Figure 5.5b), formed from well-preserved plant tissue and it forms in very poorly drained conditions, which are common in Finnish forests (Bridges, 1997).

The soil in south Finland is generally high in organic matter (Figure 5.5b), however the southern area near the coasts, which comprises nearly 26% of the nation's cultivated land, is clay rich (Figure 5.5c) with an average pH greater than 5.6 (Figure 5.5c).

5.1.1.4 Agricultural practices

A detailed overview of the Finish agricultural system is required, as it is important to explain the behaviour of ^{137}Cs in the end products considered.

For arable crops the sowing period covers the whole of May and fields are ploughed in autumn. The pasture season for dairy cows normally lasts from 10th of May to 20th of September, except for sub-areas AGR2, AGR6, AGR7, AGR8, AGR12 and AGR17 (Figure 5.3), where it lasts from the 15th of May to the 15th of September. The change from indoor feeding to grazing and vice versa is gradual, lasting a couple of weeks. Beef cattle usually do not graze, and only 3-4% of beef originates from cows that are managed as dairy cows. The forage of both dairy and beef cattle varies seasonally (Table 5.3 and Table 5.4)

The main feeds for cattle are hay, pasture and silage supplemented with other fodder plants such as kale, leaves of sugar beet, marrow kale (*Brassica oleracea*) and potatoes. For silage, different hay plants (timothy, meadow fescue etc) and to a minor extent clovers (*Trifolium sp*) are used.

During grazing in May-June, dairy cows need additional fodder grain or concentrates, approximately 0.5 kg per each kg of milk exceeding a 20 kg daily production. In late summer the same additional feed is needed for daily production exceeding 15 kg. The main additional feed of dairy cows during grazing is concentrate, including fodder grain, by-products of the food industry, molasses from sugar beet pulp, wheat bran, etc. The minimum daily portion of concentrate is 0.5 kg. Hay is given during a gradual change from indoor feeding to grazing and vice versa, however it is also given (several kg per day) during dry periods in summer when the pastures do not produce enough grass. During the period between the 7th and 26th of May 1986, about 1% of dairy cows were fed new grass. Cows did not graze, but were fed fresh grass indoors. Beef cattle usually do not graze, but the seasonal feeding of fresh grass reduces the use of other feeds. Industrial feeds for cattle are processed mainly from fodder grain coming from each agricultural production area. The grain mixtures used for cattle contain at least one-third barley or oats.

Feeding for pigs does not have a seasonal variation, as the feeds are mainly mixtures available at regional fodder factories. They consist generally of domestic cereals (barley, wheat and oat), domestic or imported meat meal and bone meal (1-5%) and marine fish meal (2-3%), and the by-products of the sugar industry (1-5%). Since the beginning of 1990, imported soya protein has largely replaced fish meal. Portions of feed constituents are varied in order to achieve an optimal raw protein concentration. Altogether, of the feeds utilized for pigs above 30% are the feed mixture described previously, 59% are fodder grains (mainly barley), and 10% are protein concentrates. In 1986 some 20% of the pork consumed originated from farms where whey is given to pigs. It contributes to the protein and carbohydrate fractions of the feed. The use of whey decreased substantially towards the early 1990s.

The harvest of cereals starts in the last week of July (rye) and lasts until the first three weeks of September. The growth period is 97-106 days for spring wheat, 95-101 days for oats, and 84-100 days for barley, varying with varieties sown. Table 5.4 reports the cereal yield for each agricultural region.

Grasslands are cultivated either for hay, pasture or silage. The main plant species used are timothy (*Phleum pratense*), meadow grass (*Festuce pratensis*) and clovers such as *Trifolium pratense*, *Trifolium hybridum* and *Trifolium repens*. Depending on the plant mixture, grasslands with a cover stand are ploughed every 2 or 3 years, with a timothy stand every 3 to 4 years and with meadow grass every 3 to 5 years.

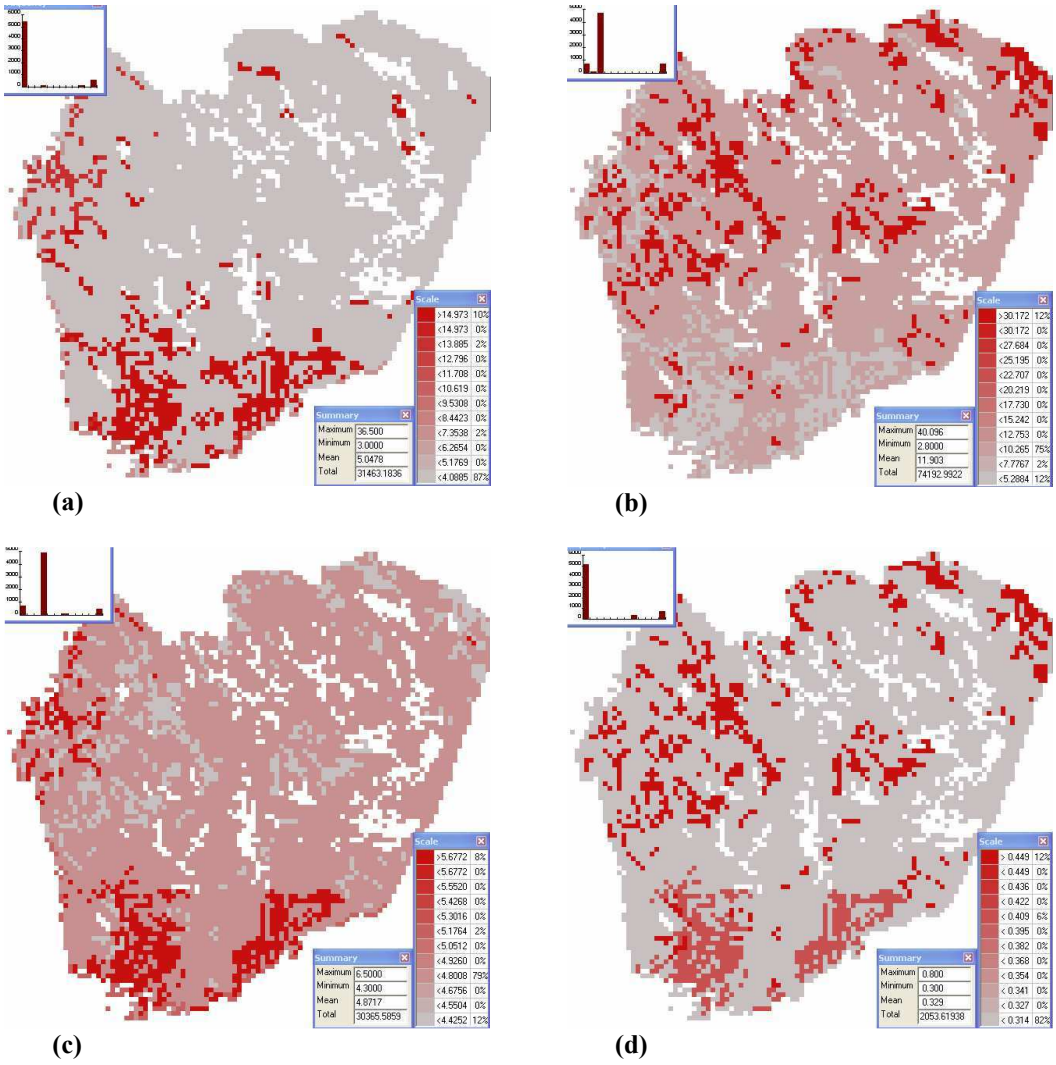


Figure 5.5 South Finland soil characteristics: clay (a), organic matter (b), pH (c) and exchangeable K^+ (d). The soil maps have been obtained from the SAVE-IT software, which are derived from the European Soil Data Base (EC, 1995).

Table 5.3 Total amount of feed supplied to beef cattle and dairy cows in 1988. Although the table shows measurements for 1988, they can be regarded as typical values (IAEA, 1996).

	Feed (kg)							
	Area	Silage	Hay	Pasture	Feed grain	Complete feed	Concentrate	Others
Beef cattle	AGRl	1474.2	847	1209	1008.7	78.1	41.8	196
	AGR2	2154.6	466.4	2853.5	727.1	146.3	33	520.8
	AGR3	812.7	761.2	604.5	1057.1	198	62.7	778.4
	AGR4	1864.8	822.8	1326	903.1	86.9	69.3	442.4
	AGR5	1474.2	686.4	2359.5	919.6	103.4	55	347.2
	AGR6	2627.1	517	1904.5	689.7	215.6	49.5	280
	AGR7	2186.1	596.2	2788.5	706.2	161.7	33	95.2
	AGR8	2916.9	591.8	2379	612.7	209	36.3	67.2
	AGR9	1083.6	946	1079	1049.4	47.3	61.6	224
	AGR10	1997.1	600.6	2411.5	716.1	177.1	35.2	168
	AGR11	-	-	-	-	-	-	-
	AGR12	1354.5	891	1339	937.2	92.4	67.1	212.8
	AGR13	1927.8	699.6	2411.5	751.3	154	47.3	168
	AGR14	1436.4	776.6	1612	882.2	183.7	40.7	291.2
	AGR15	1430.1	908.6	1527.5	1050.5	23.1	84.7	179.2
	AGR16	1266.3	772.2	1547	1003.2	71.5	83.6	425.6
	AGR17	1222.2	576.4	2067	875.6	94.6	88	-
Dairy cows	AGR1	5657.4	1240.8	6370	1312.3	210.1	226.6	711.2
	AGR2	7188.3	882.2	6123	938.3	552.2	187	604.8
	AGR3	6615	1322.2	3900	1118.7	833.8	319	1080.8
	AGR4	5770.8	1276	5791.5	1294.7	292.6	259.6	968.8
	AGR5	6375.6	1018.6	6526	1133	342.1	192.5	1024.8
	AGR6	7515.9	827.2	5453.5	866.8	814	191.4	459.2
	AGR7	6835.5	1080.2	6389.5	779.9	690.8	170.5	397.6
	AGR8	8246.7	869	6513	830.5	570.9	203.5	229.6
	AGR9	5242.6	1425.6	5908.5	1277.1	206.8	202.4	482.6
	AGR10	6967.8	1007.6	6448	964.7	497.2	272.7	403.2
	AGR11	5550.3	1533.4	4504.5	1665.4	139.7	244.2	392
	AGR12	6161.4	1226.6	6272.5	1229.8	325.6	221.1	492.8
	AGR13	6923.7	1104.4	6363.5	995.5	438.9	199.1	308
	AGR14	5707.8	1262.8	6181.5	990	514.8	166.1	509.6
	AGR15	5663.7	1331	5408	1501.5	206.8	221.1	620.4
	AGR16	4775.4	1460.8	5304	1375	264	300.3	873.6
	AGR17	6463.8	1311.2	3672.5	1221	570.9	272.8	548.8

Table 5.4 Cereals yield in 1986, measurements expressed in millions of kg (IAEA, 1996).

	Cereal yield (10^6 kg)				
Area	**Winter wheat**	**Spring wheat**	**Rye**	**Barley**	**Oats**
AGR1	-	6	1.3	41.5	42.1
AGR2	0.3	8.6	6.2	215.7	206.6
AGR3	3.6	28.4	2.5	16.3	11.7
AGR4	4.4	40.3	5.8	126.2	96.9
AGR5	0.3	11.5	2.1	50.2	40.7
AGR6	-	-	1.1	57.4	17.9
AGR7	0.2	1	2.5	51.9	51.1
AGR8	-	2	3	88.1	37.3
AGR9	1	13.7	2.2	47.1	47.5
AGR10	-	4.4	3.2	43.5	47.5
AGR11	2.2	68.1	9.2	71.8	31.5
AGR12	2.7	6.5	3.5	65.1	59.3
AGR13	-	2.1	1.8	53.5	41.2
AGR14	5.6	29.1	5.6	136.3	141.2
AGR15	2.5	62	3.9	120.8	65.6
AGR16	32.5	177.9	13.3	251.1	106.4
AGR17	-	11	1.4	124.5	82.3

5.1.1.5 ^{137}Cs activity concentration in food products.

The four end products used to test the model predictions are dairy milk, beef, pork and cereals (i.e. wheat).

In 1986 the mean ^{137}Cs concentration in wheat was 4.9 Bq/kg, with a rather high standard error, ranging between 3.4 and 7.4 Bq/kg (Figure 5.6). This level of variability is mostly due to the uneven deposition over the tested area (Figure 5.2). By 1987 the mean radiocaesium concentration dropped approximately one order of magnitude, 0.53 Bq/kg, thereafter the annual contamination shows a marginal reduction (Figure 5.6). The reduction in wheat ^{137}Cs activity concentration, between 1986 – 1987, could be accounted for the countermeasures strategy implemented in the region, nevertheless the documentation (IAEA, 1996) does not provide a detailed overview of the countermeasures adopted, while the slow decline between 1987 – 1990 is predominant due to the soil type (fine mineral soil) which tends to fix radiocaesium and reduce its bioavailability.

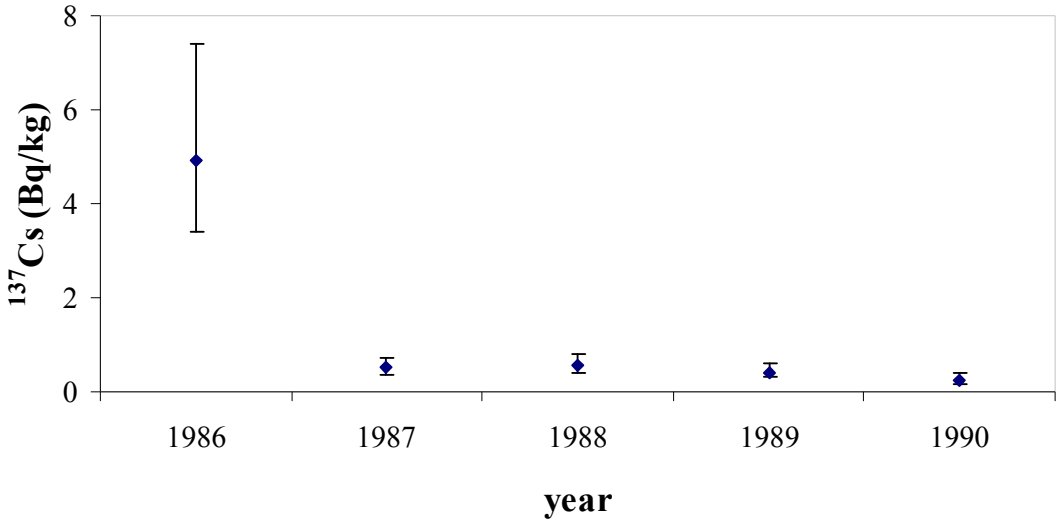


Figure 5.6 ^{137}Cs activity concentration in wheat for the south Finland area. The Error bars represent the 95% confidence interval.

The ^{137}Cs activity concentration in dairy milk did not rise immediately after the initial deposition in 1986, as the dairy cows were still on a stored feed diet; the mean concentration was 2 Bq/l in May 1986 (Figure 5.7). However it rose to 30 Bq/l in June as the cows were transferred to pasture. The decline in milk contamination during the 1986 summer is mainly due to the weathering of radiocaesium from the pasture surface. Nevertheless a significant increase is observable in the last few months of the same year (Figure 5.7). This trend could be accounted for the integration of hay and silage, which were exposed to the radiocaesium for the entire summer into the cow's diet. In the following years, radiocaesium concentration in milk shows a progressive decline, particularly during summer due to the cows feeding regime, which passes from stored feed to pasture. The rate of decline of milk activity concentration is relatively slow. The ^{137}Cs activity concentration at the end of 1990 is still higher than the concentration in May 1986 at about 10% of the peak concentration recorded in June 1986.

The observed mean ^{137}Cs activity concentration for beef (Figure 5.8) exhibits a similar trend to dairy milk, which is due to the similar feeding regime (Table 5.3). However, the cattle contamination is generally a factor of four higher than the milk, which is due to the higher transfer factor (IAEA, 1994).

The concentration increases rapidly from 9.2 to 134 Bq/kg in the first seven months, before it slowly declines and by end of 1990 the concentration is comparable to the measurements recorded in May 1986 (Figure 5.8).

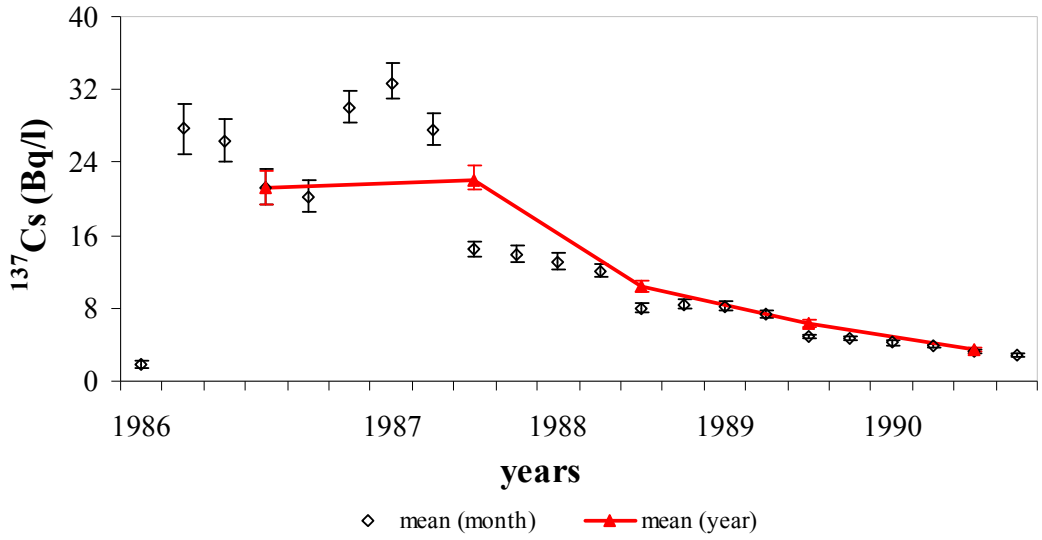


Figure 5.7 ^{137}Cs activity concentration in dairy milk. The points represent monthly measurements, while the solid line is the year mean. Errors bars refer to 95% confidence interval.

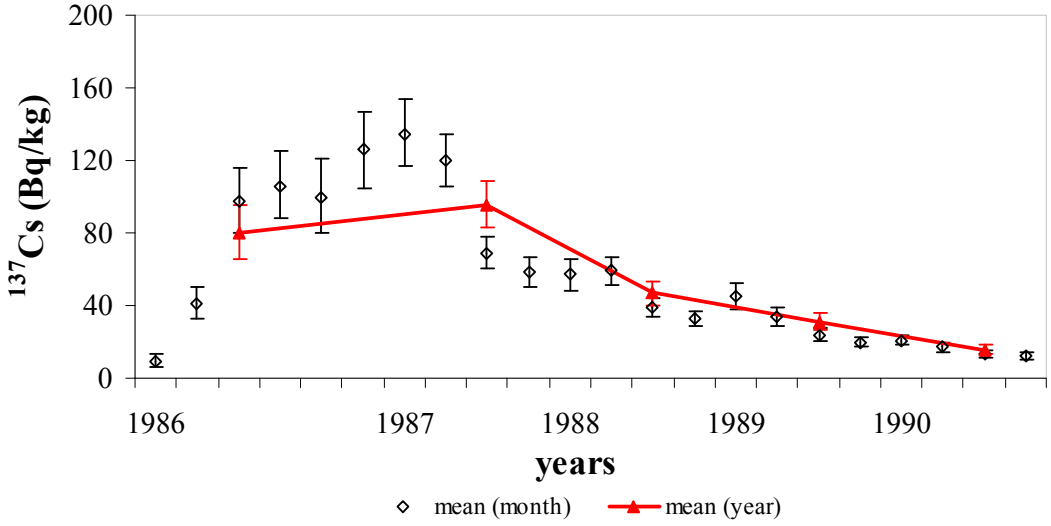


Figure 5.8 ^{137}Cs activity concentration in beef. The points represent monthly measurements, while the solid line is the yearly mean. Errors bars are the 95% confidence interval.

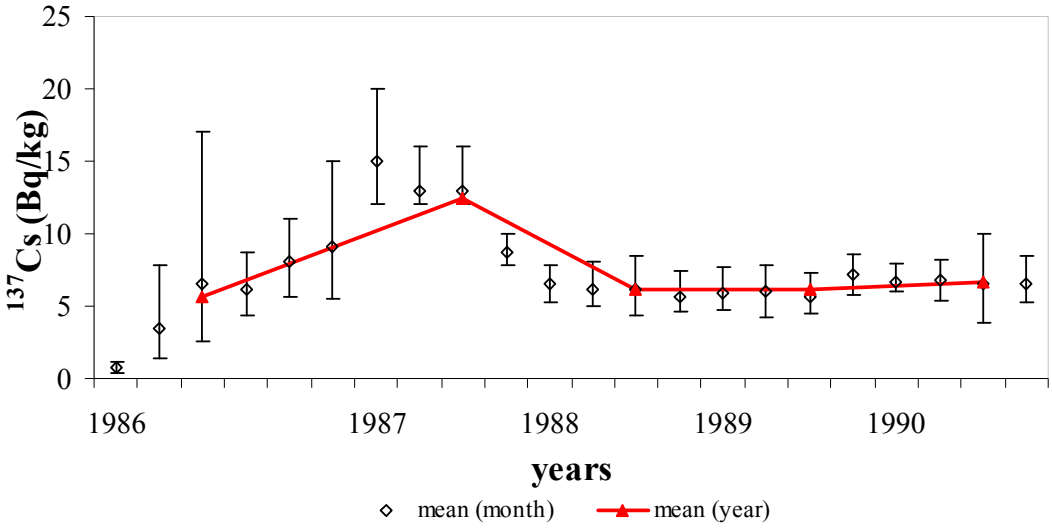


Figure 5.9 ^{137}Cs activity concentration in Pork. The points represent monthly measurements, while the solid line is the yearly mean. Errors bars refer to 95% confidence interval.

The observed radiocaesium activity concentration in pork is 0.8 Bq/kg in May 1986 and peaks at about 15 Bq/kg during the summer of 1987. From 1987 to 1990 the concentration remains between 6 to 7 Bq/kg (Figure 5.9).

The ^{137}Cs concentration in pork is characterised by a trend which is similar to that observed for milk and beef, however the peak concentration is 10 to 12 times lower than the one for beef. In contrast to the ^{137}Cs concentrations in the other two food products, there is not a significant decrease in the ^{137}Cs activity concentration of pork in the period between 1987 and 1990 (Figure 5.9). This is probably due to the diet which is dominated by grain.

5.1.2 Semi-natural scenario: Bad Waldsee state forest, Germany

Model predictions have been compared with 5 years measurements (1987 – 1992) of ^{137}Cs activity concentration in roe deer (Bq/kg) from the state forest of Bad Waldsee (Figure 5.10) which is situated 30 km from Lake Constance, near the town of Ravensburg in southern Germany. The site received 27 kBq/m^2 of ^{137}Cs (Figure 5.11) (Zibold personal communication) on the 1st of May 1986, mostly as wet deposition (Figure 5.12) (Zibold et al 2001).

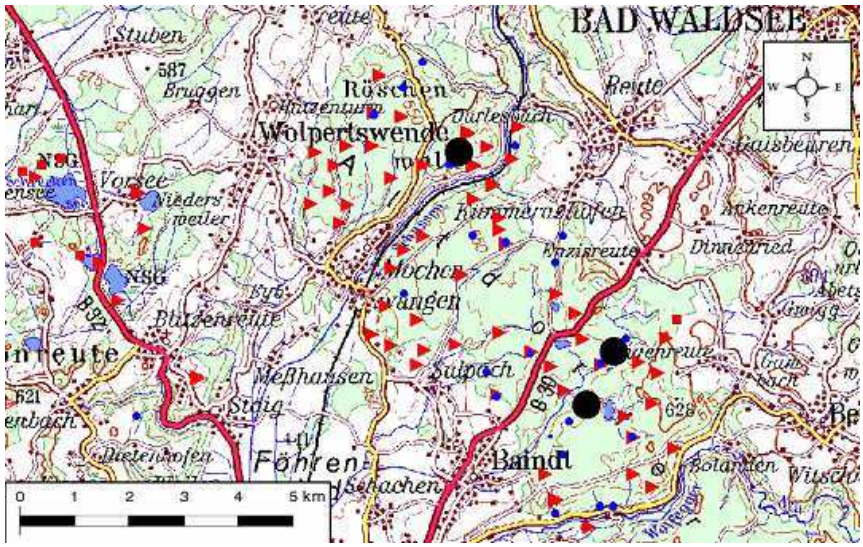


Figure 5.10 Bad Waldsee state forest. Blue dots represent the soil sampling sites in 1987; Red triangles: sites of roe deer shooting in 1987; Red squares: sites of roe deer shooting in 2003; black dots from top: soil profile, *Xerocomus badius* and spruce tree sampling (Zibold et al 2001).

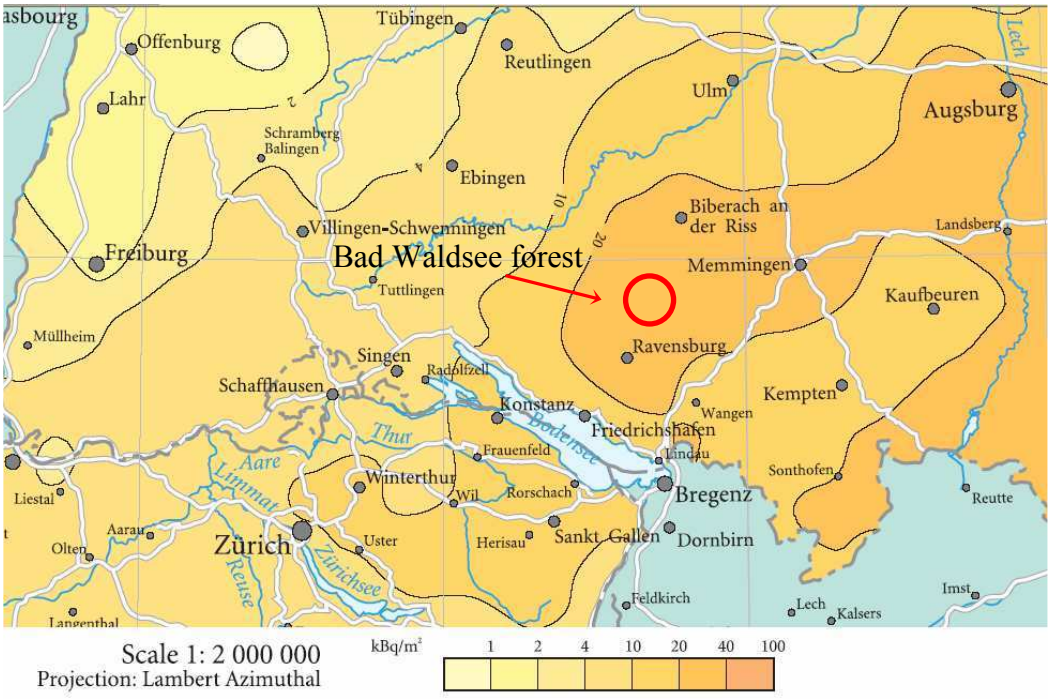


Figure 5.11 ^{137}Cs deposition on South of Germany, 1986. (De Cort, 2001)

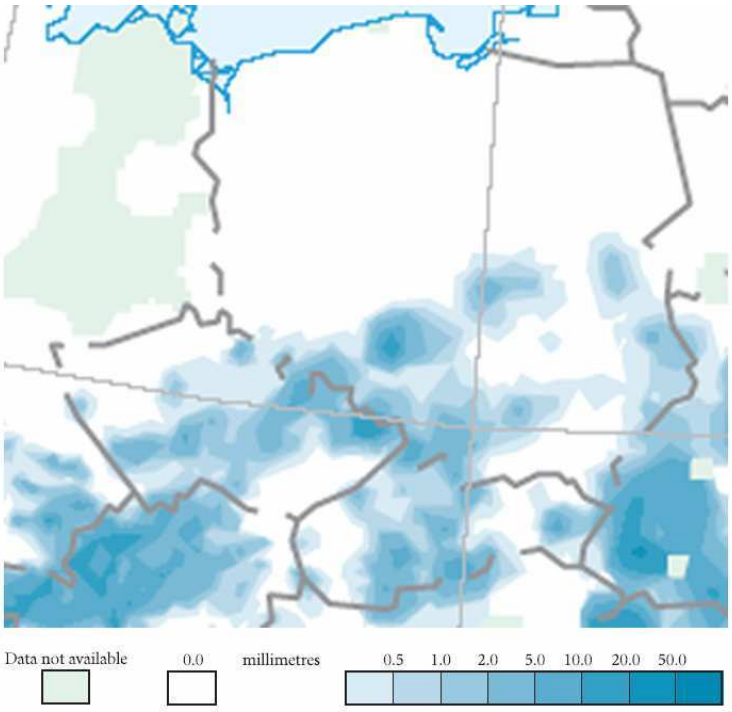


Figure 5.12 Precipitation measurements for the 1st of May 1986 on Germany. (DE CORT, 2001).

The forest comprises 41 km^2 of hunting area, 30 km^2 of forest, which is mainly spruce, (*Picea abies*) however there is a small area of mixed forest (beech, *Fagus* and spruce, *Picea abies*), 10 km^2 of agricultural land and 0.9 km^2 of fresh water (Zibold personal communication).

The mean annual temperature is 8˚C and the Wetterwarte Bad Schussenried reports, for the years 1980 – 1994, an average maximum temperature of 17.8˚C in July and a minimum value of –1.3˚C in January. Maximum precipitation is recorded in July, average value 123 mm, and the minimum 44 mm in February for a total of 905 mm per year. (Zibold et al, 2001)

5.1.2.1 Soil characteristics

The main type of soil is *Luvisol*, which is a common type of soil for the entire European region and the geology of the bedrock is mainly *Moraine* (Zibold et al 2001). The site is characterised by an average clay content of 21% (Figure 5.13a), and very low organic matter, (i.e. < 4%) except for few localised areas where the organic matter is higher than 19%, (Figure 5.13b). Consequently it is an acid environment, with an average pH of 6.6, (Figure 5.13c). The exchangeable potassium

is low in the entire site, (Figure 5.13d) therefore a high plant uptake of radiocaesium present in the soil was expected.

Soil sampling was performed from 1986 to 2000 in order to identify the average radiocaesium distribution in the soil profile; 209 soil samples were collected from the whole forest area (Zibold personal communication). The ^{137}Cs is mostly located (i.e. > 90%) in the top 10 cm of the soil profile (Figure 5.14), even fourteen years after the initial deposition (Zibold et al, 2001).

The sampling and analysis methodology has been extensively described in the literature (Zibold et al, 2001; Kilefer et al, 1996; Lindner et al, 1994) and therefore will not be discussed in the present work.

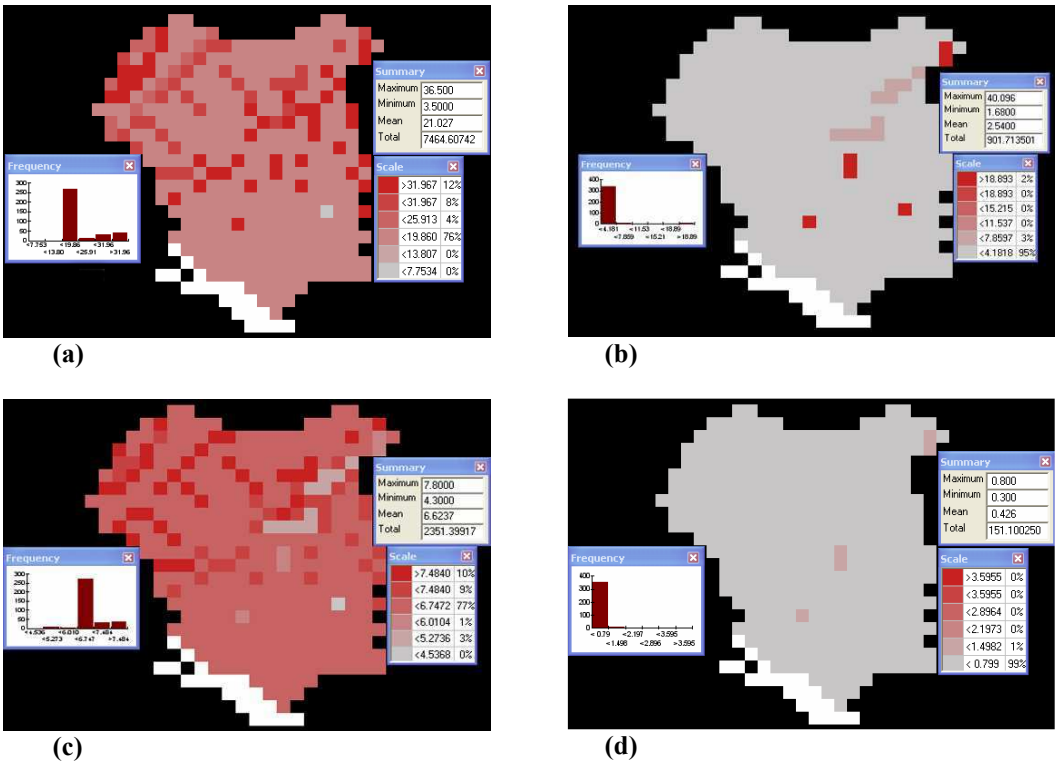


Figure 5.13 South Germany soil characteristics: clay (a), organic matter (b), pH (c) and exchangeable K^+ (d). The soil maps have been obtained from the SAVE-IT software, which are derived from the European Soil Data Base (EC, 1995).

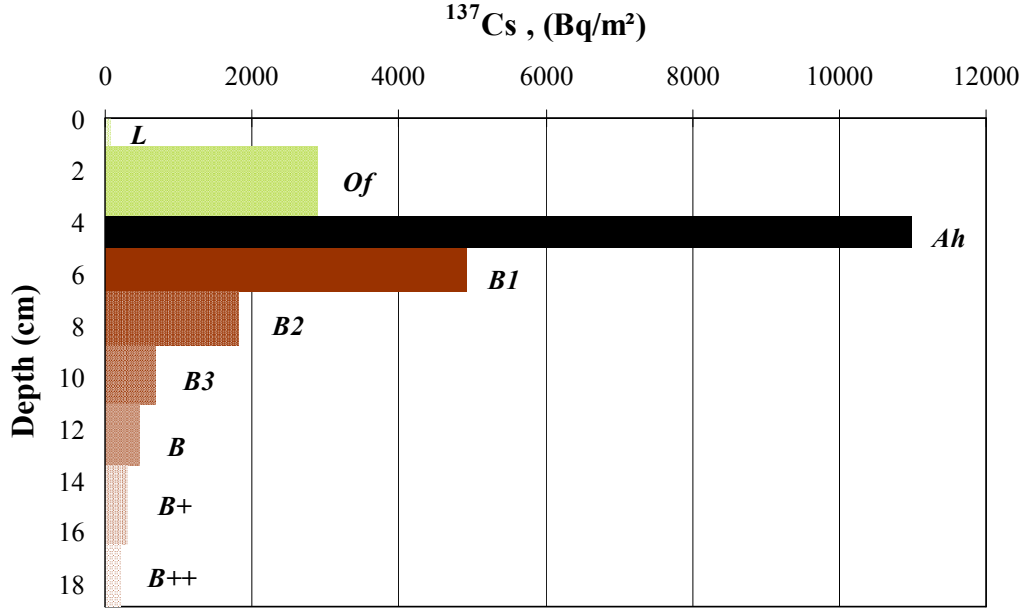


Figure 5.14 Depth distribution of ^{137}Cs in soil mixed forest site. (Zibold personal communication).

5.1.2.2 Tree characteristics

The dominant tree species in the region is spruce, *Picea abies*, comprising 80% of the total forest area (Zibold et al, 2001). The tree density has been estimated to be 347 trees per hectare (Zibold personal communication). The average age at the time of contamination was 56 years. The average height is 33.6 m while the trunk diameter is 38 cm at the height of 1.5 m (Zibold et al, 2001).

5.1.2.3 Forest game.

The main game species are roe deer, *Capreolus capreolus,* 30 – 40 animals per km² (Zibold et al, 2001) and transient wild boar, *Sus scrofa scrofa L.* Game graze principally in the forest area, however they can occasionally be found in the agricultural sites, although they have a minor effect on the diet.

The hunting season for male roe deer lasts five months, from the 15th of May to the 15th of October. Adult females with their fawns, the hunting season begins on the 1st of September and ends at the end of January for adult females, while for fawns it ends in February (Zibold et al, 2001).

The roe deer contamination has been monitored since 1987. During this period more than 7000 animals have been shot (Zibold et al, 2001) and 10 – 100 g of fresh muscle have been analysed in each occasion (Lindner et al, 1994).

Between 1987 – 1991 the animals were hunted randomly on the whole forest area, however in the following years the sampling approach changed and roe deer were shot only in selected forest areas where animals had been recorded to have a higher ^{137}Cs activity concentration (Zibold et al, 2001). Therefore only the sample for the period 1987 – 1991 can be regarded as a representative measurement of the overall forest contamination.

The average weight of a living individual is about 17 kg, including 1 kg head and 1 kg blood. More than 50% of the roe deer shot were younger than 1 year (Zibold et al, 2001). The ^{137}Cs activity concentrations are log-normal distributed with mean value of 310.87 Bq/kg (fw) for the period 1987 to 1991.

The results reveal a general seasonal variation in roe deer contamination, (Figure 5.15). Similar patterns have been observed in comparable studies in several parts of Europe, for instance in Sweden (Karlen et al, 1991) and it has been related to the mushroom season. It is well known that mushrooms are part of the roe deer diet. In addition, the chestnut boletus (*Xeromocus badius*), which is very abundant in the region, can accumulate Cs to a considerable extent and therefore contribute considerably to the seasonal variability of the roe deer contamination (Zibold personal communication).

The annual ^{137}Cs activity concentration (Figure 5.16) shows a slow exponential decrease with a half-life of 2.6 ± 0.4 years. The 95% confidence interval on the estimated annual mean is rather low.

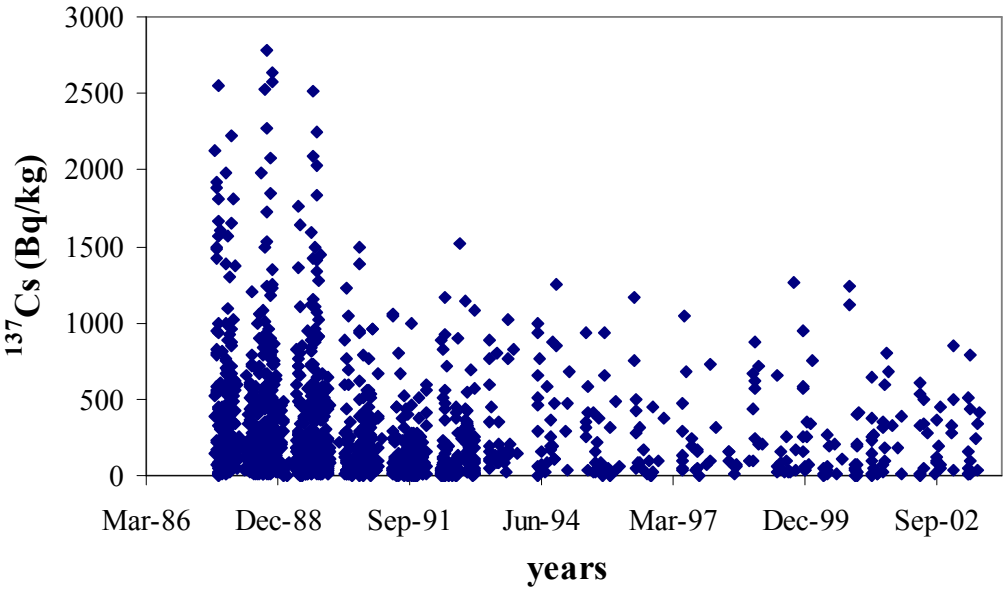


Figure 5.15 Single values of ^{137}Cs activity concentration in roe deer (Bq/kg).

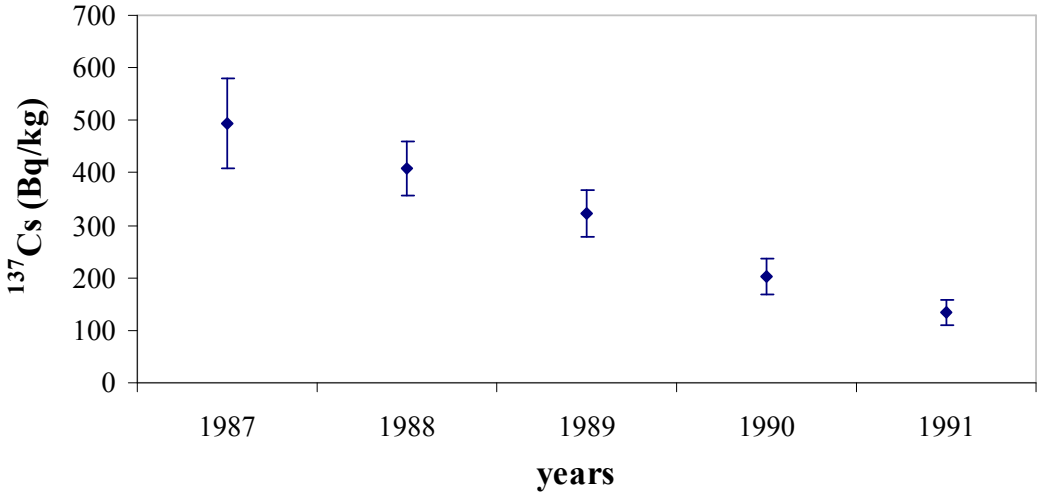


Figure 5.16 The annual average of ^{137}Cs activity concentration in roe deer for the five years period (1987 – 1991) is reported (♦) with the 95% confidence interval on the mean.

5.2 METHOD

The accuracy of the model prediction has been quantified using the mean of the relative error between the model prediction (x) and the empirical data (y), as it is an intuitive measurement of the relative error of the model predictions.

$$MRE = \frac{\sum_{i=n}^{1} \frac{|x_i - y_i|}{y_i}}{n} \tag{4.1}$$

where

x is the model predictions

y is the empirical data

n is the total number of x-y couplets

5.3 MODEL – OBSERVATIONS COMPARISON OF RESULTS

The radiocaesium activity concentrations in the considered end-products have been measured on a monthly basis, nevertheless the annual mean has been used for the model-scenario test as RIFE1 and TEMAS rural and semi-natural model are characterised by an annual time resolution.

5.3.1 Rural scenario: South Finland

5.3.1.1 Dairy milk (Bq/kg)

The two models (SAVE and TEMAS) show different prediction trends. Initially, SAVE overestimates the radiocaesium activity concentration in dairy milk (Bq/kg) relative error 0.77 (Figure 5.17, Table 5.5). In the following three years the model prediction appears to be lower than the observations, nevertheless, in the final year, 1990, the model shows a high level of accuracy, relative error 0.04. Overall the model estimates an exponential decrease in ^{137}Cs concentration in dairy milk which is in agreement with the observations trend.

TEMAS highly underestimates the dairy milk contamination for the year 1986 and 1987 (Figure 5.18). Nevertheless between 1988 and 1990 the prediction accuracy

gradually increases; in the final year the relative error is 0.26 (Figure 5.18, Table 5.5). However, the model prediction trend does not appear to resemble the observation pattern; except for the first year, where a significant decline in the milk contamination is present, the model predicts a very slow decrease rate, which implies that TEMAS would largely overestimate the ^{137}Cs activity concentration in dairy milk in the long term.

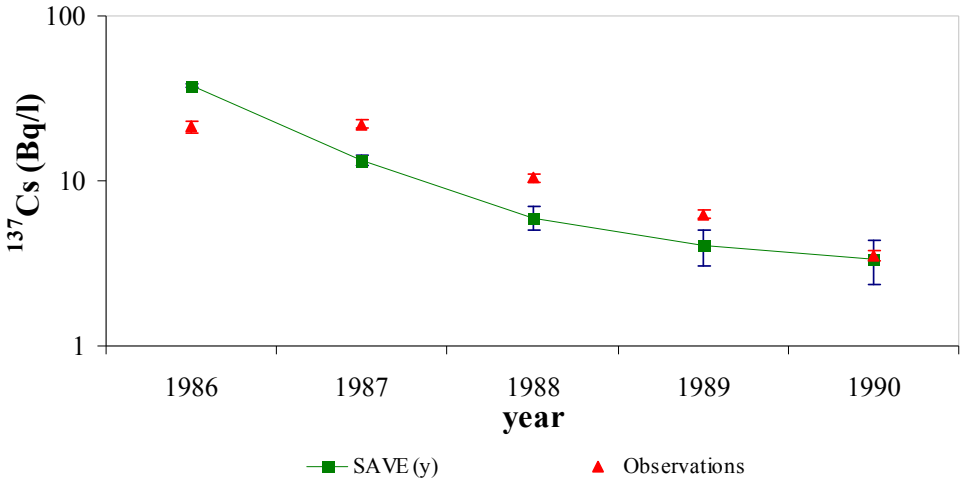


Figure 5.17 Dairy milk (Bq/l): SAVE rural (■) predictions are compared with the observations (▲), vertical bars indicate 95% confidence intervals on the mean value of observations.

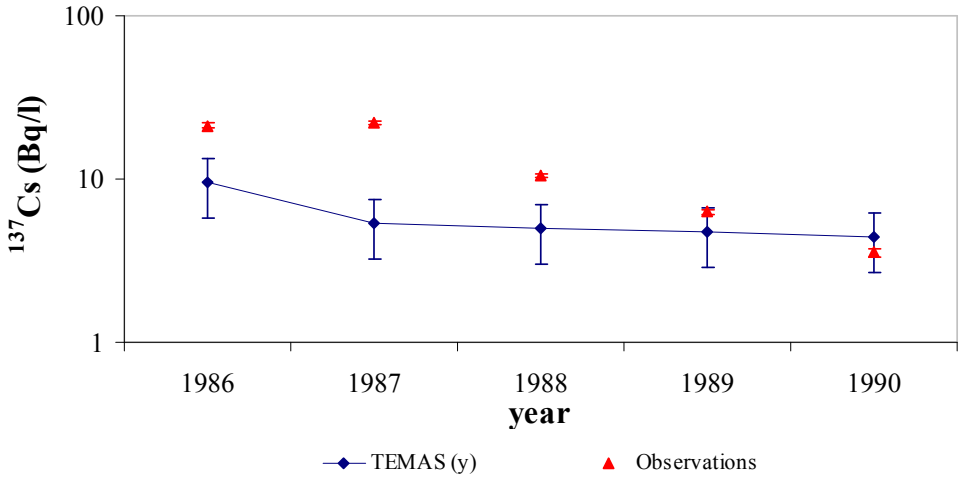


Figure 5.18 Dairy milk (Bq/l): TEMAS rural (♦) predictions are compared with the observations (▲), vertical bars indicate 95% confidence intervals on the mean value of observations.

5.3.1.2 Cattle beef (Bq/kg)

The prediction trends for beef contamination by radiocaesium (Bq/kg) resemble the one presented for dairy milk, section 4.3.1.1. In the first year, i.e. 1986, SAVE overestimates the ^{137}Cs concentration in beef relative error 1.13, however in the following three years the predicted contamination is higher than the observations, relative error ranging between 0.54 and 0.45. Nevertheless the estimation of 1990 is highly accurate, relative error 0.13 (Figure 5.19, Table 5.5).

TEMAS provides an estimation which is higher than the 95% confidence interval on the observation value for the years 1986 to 1989, relative error 0.58 and 0.72 respectively. However the final prediction shows a significant level of accuracy, relative error 0.25 (Figure 5.19, Table 5.5). In spite of this, the time dynamic predicted by TEMAS, for the period 1987 – 1990, is lower than the rate observed for the measurements, therefore in the long term the model might tend to overestimate the beef contamination.

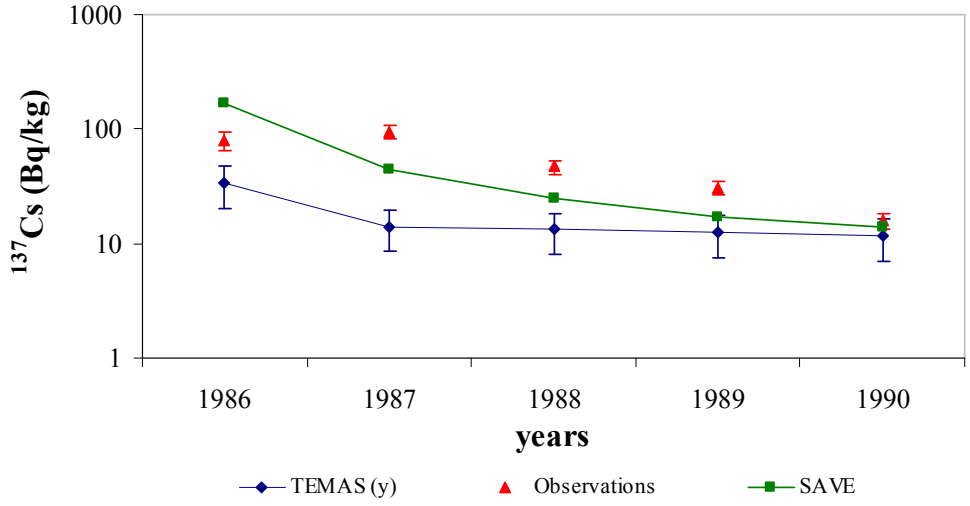


Figure 5.19 Beef (Bq/l): SAVE rural (■) and TEMAS rural (♦) predictions are compared with the observations (▲), vertical bars indicate 95% confidence intervals on the mean value of observations.

5.3.1.3 Pork (Bq/kg)

SAVE generally overestimates the ^{137}Cs activity concentration in pork (Bq/kg) for the entire period, i.e. 1986 – 1990, relative error 0.70 and 0.55 respectively (Table 5.5). Nevertheless the prediction pattern shows similarities with the observations. A concentration build up is estimated within the first year of

deposition. A decline is observable between 1987 and 1990 which is followed by a relative equilibrium for the period 1988 – 1990, (Figure 5.20).

TEMAS provides an accurate estimation of pork contamination, however the trend for the first 2 years does not resemble the measurement, as the observations show a concentration increase, while the model predicts a sharp decrease, between 1986 and 1987.

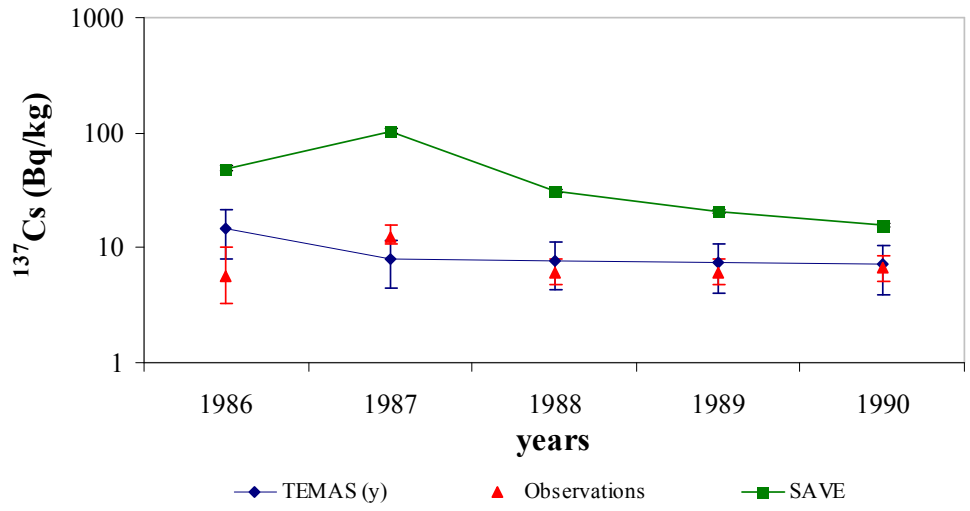


Figure 5.20 Pork (Bq/l): SAVE rural (■) and TEMAS rural (♦) predictions are compared with the observations (▲), vertical bars indicate 95% confidence intervals on the mean value of observations.

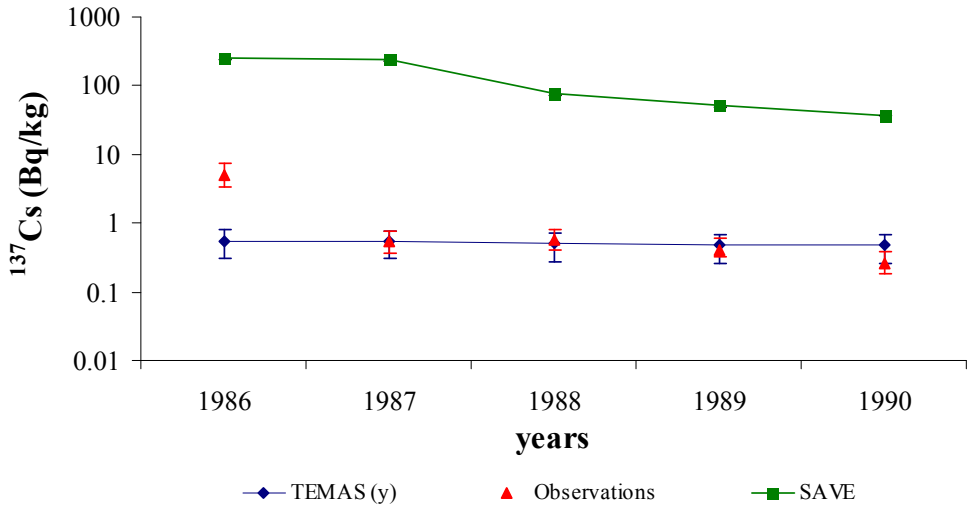


Figure 5.21 Cereals (Bq/l): SAVE rural (■) and TEMAS rural (♦) predictions are compared with the observations (▲), vertical bars indicate 95% confidence intervals on the mean value of observations.

5.3.1.4 Cereals (Bq/kg)

SAVE overestimates the radiocaesium activity concentration in cereals (Bq/kg) by two orders of magnitude for the entire period considered. TEMAS underestimates the radiocaesium concentration in the first year of deposition, relative error 0.89. However, from 1987 to 1990 the TEMAS predictions and the observations have a high level of agreement. Nevertheless, in the long term the model might overestimate the cereal contamination, as the rate of decline is much lower than that observed. The predicted half-time for the 1987 – 1990 period is 10.8 years, while the measurements suggest a half-time of 1.75 years which is similar to that predicted by SAVE.

5.3.2 Semi-natural scenario: Bad Waldsee

SAVE overestimates the ^{137}Cs activity concentration in roe deer for the whole studied period, relative error ranging between 0.56 and 3.03. The model prediction shows a build-up in the first year, i.e. 1987 – 1988, in the following years a decrease is predicted (half-time 3.9 years) which is similar to the half-time of 3.7 years estimated from the measurements (Figure 5.22).

RIFE1 generally overestimates the roe deer contamination, relative error ranging between 1.22 and 7.44. The overall trend has a similar pattern to the SAVE predictions, i.e. build-up in the first years followed by a decrease (half-time is 21.5 years), however the reduction rate for the period 1988 – 1990 is generally lower than SAVE and the observations.

The ^{137}Cs activity concentration in roe deer estimated by FORM shows the opposite trend to SAVE, RIFE1 and the data. The model predicts a gradual increase of radiocaesium concentration in the roe deer meet (Figure 5.23).

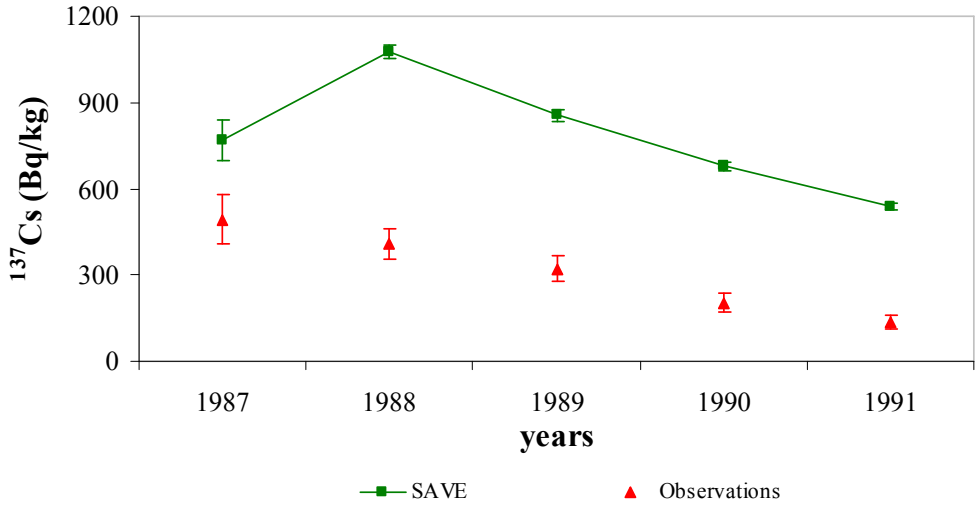


Figure 5.22 The SAVE prediction of roe deer contamination (■) is compared with the observations (▲). The error bars are the 95% confidence interval on the mean value.

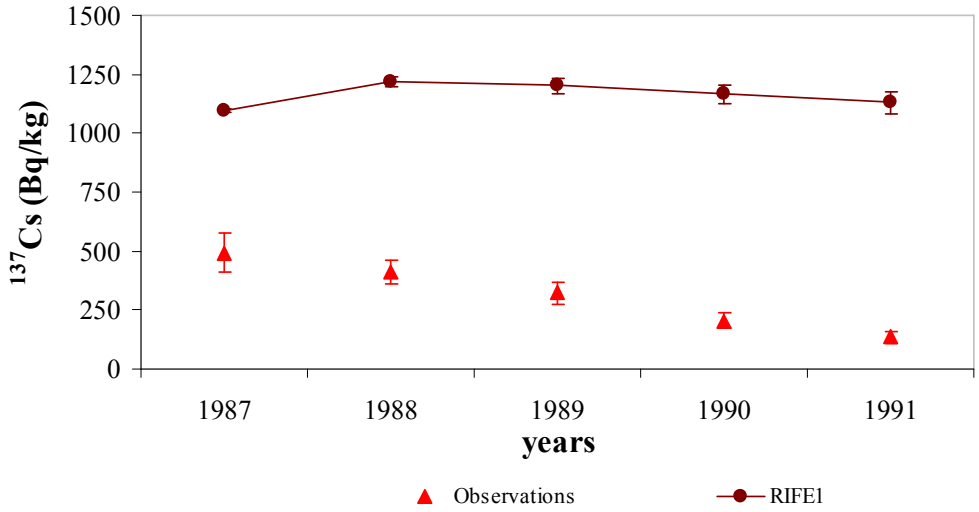


Figure 5.23 The RIFE1 prediction of roe deer contamination (●) is compared with the empirical data (▲). The error bars represent the 95% confidence interval.

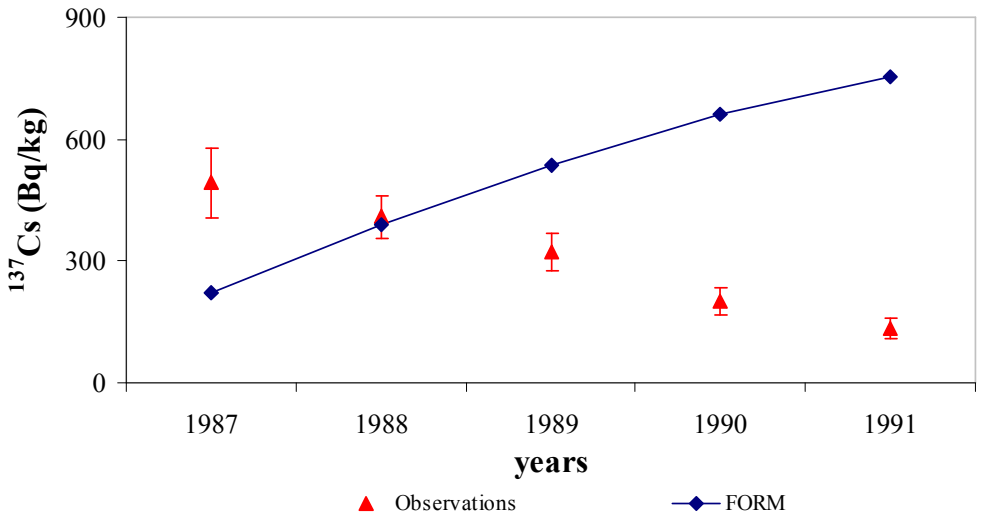


Figure 5.24 The FORM (TEMAS semi-natural) prediction of radiocaesium activity concentration (Bq/kg) in roe deer (♦) is compared with the observations (▲). The error bars represent the 95% confidence interval.

5.4 DISCUSSION

5.4.1 Rural scenario

A good agreement between the SAVE prediction of dairy milk (Bq/l) and beef (Bq/kg) radiocaesium concentration has been observed (Figure 5.25 and 5.26), MRE equal to 0.36 and 0.54 respectively (Table 5.5). In addition, comparing the model estimates with the measurements on a monthly time step, it is possible to observe that the SAVE predictions are subject to the seasonal variation of the contaminant (Figure 5.27 and 5.28). This is due to the flexible user interface which allows the user to customise the animal management to be as realistic as possible.

On the other hand the model tends to overestimate the activity concentration in pork (Bq/kg) and cereals (Bq/kg) by several orders of magnitude (Figure 5.29 and 5.30).

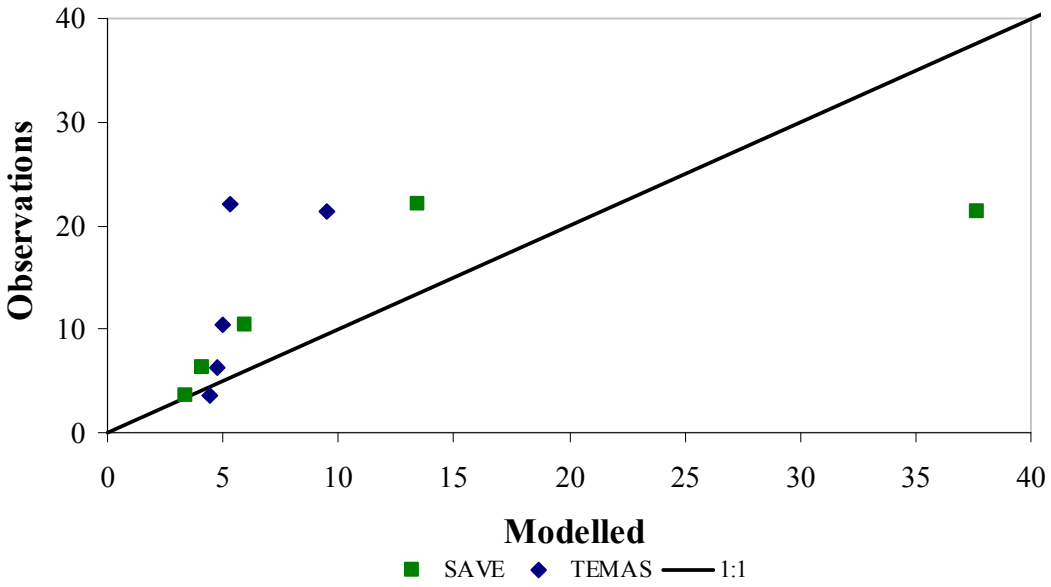


Figure 5.25 The SAVE rural (■) and TEMAS rural (♦) model prediction for dairy milk are compared with the observations, the solid line represents the 1 to 1 relationship.

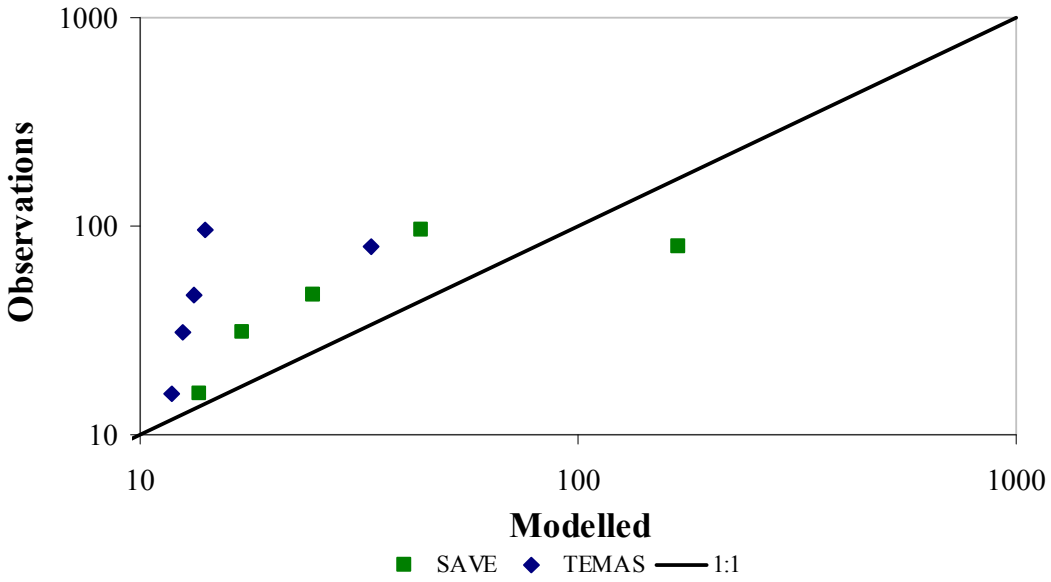


Figure 5.26 The SAVE rural (■) and TEMAS rural (♦) model prediction for cattle beef are compared with the observations, the solid line represents the 1 to 1 relationship.

Table 5.5 The relative error between models prediction for rural food products and the observation is reported for each year investigated. In addition the mean relative error (MRE) is shown at the bottom of the table.

	Relative Error							
	Dairy milk		Beef		Pork		Cereals	
	SAVE	TEMAS	SAVE	TEMAS	SAVE	TEMAS	SAVE	TEMAS
1986	0.77	0.55	1.13	0.58	0.70	1.56	50.62	0.89
1987	0.39	0.76	0.54	0.85	0.92	0.35	441.84	0.02
1988	0.42	0.52	0.47	0.72	0.75	0.26	131.73	0.11
1989	0.35	0.25	0.45	0.59	0.64	0.20	126.14	0.18
1990	0.04	0.26	0.13	0.25	0.55	0.08	142.95	0.82
MRE	0.40	0.47	0.54	0.60	0.71	0.49	178.66	0.40

Table 5.6 The relative error between the roe deer contamination predicted by the models and the observation is reported for each year investigated. In addition the mean relative error (MRE) is shown at the bottom of the table.

	Roe Deer Relative Error		
	SAVE	FORM	RIFE1
1986	0.56	0.55	1.22
1987	1.63	0.05	1.97
1988	1.65	0.67	2.74
1989	2.35	2.26	4.76
1990	3.03	4.63	7.44
MRE	1.84	1.63	3.63

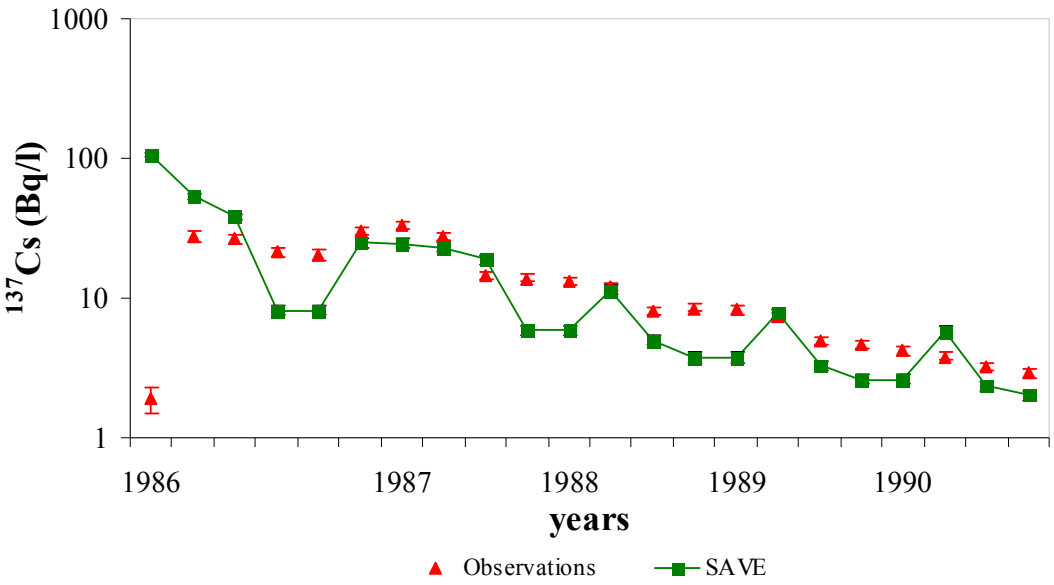


Figure 5.27 The SAVE rural model prediction of radiocaesium activity concentration in dairy milk is compared with the observation on a monthly time scale.

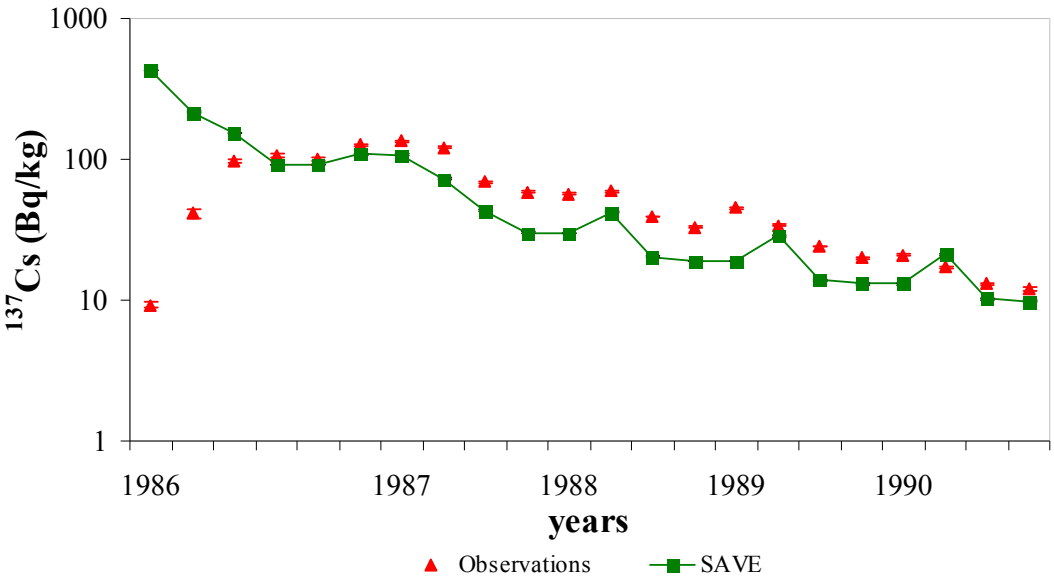


Figure 5.28 The SAVE rural model prediction of ^{137}Cs activity concentration in cattle beef is compared with the observation on a monthly time scale

The type of feed has a strong impact on the animal contamination, therefore where SAVE allows a realistic description of the feeding regime, providing the seasonal intake (kg dw/d) of pasture, stored grass, maize silage and concentrate, the system produces an accurate prediction, i.e. dairy milk and beef. On the contrary, where the user-interface is not flexible the prediction is inaccurate, as is shown by the pork prediction (Table 5.5). In SAVE, pigs are assumed to be fed only with concentrate all through the year, unfortunately the pigs diet is more complex than this, as described in the section 5.1.1.6.

The predicted ^{137}Cs activity concentration in cereals might be overestimated due to the accuracy of the model inputs and land cover spatial data implemented in the system. The SAVE spatial database is mostly based on the European soil database (EC, 1995), However the Scandinavian countries spatial data are based on the FAO soil database (FAO-AGLL, 2000), which is characterised by a lower resolution and therefore lower accuracy than the European soil database. Consequently, the cultivated areas, considered in this exercise, might have different soil characteristics than the one reported in the SAVE spatial database. In addition, the soil in the considered areas might have been improved throughout the years to increase the crop production. All this might have affected the model predictions accuracy.

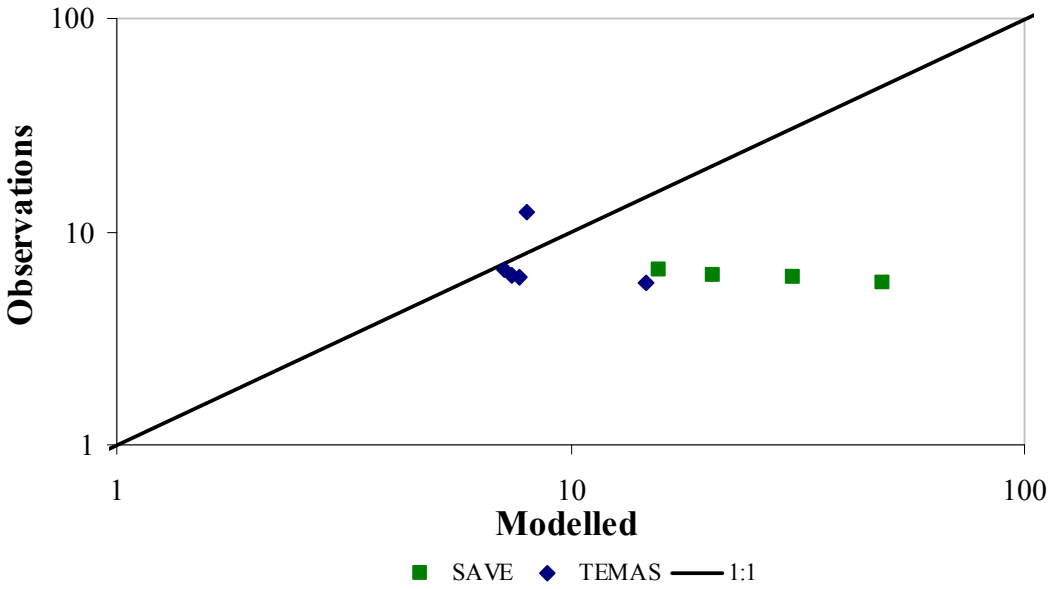


Figure 5.29 The SAVE rural (■) and TEMAS rural (♦) model prediction for pork are compared with the observations, the solid line represents the 1 to 1 relationship.

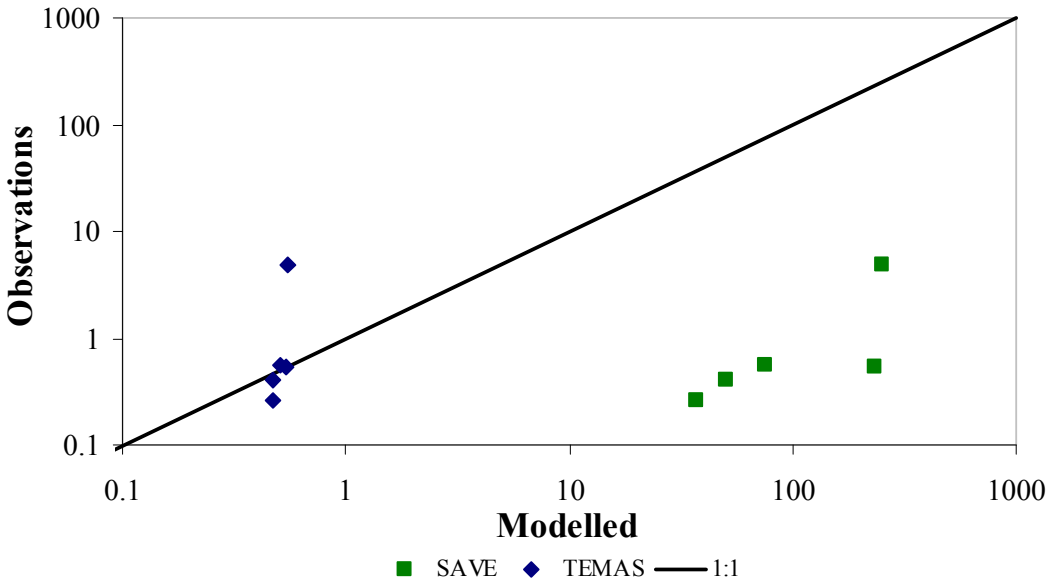


Figure 5.30 The SAVE rural (■) and TEMAS rural (♦) model prediction for ^{137}Cs activity concentration in cereals are compared with the observations, the solid line represents the 1 to 1 relationship.

TEMAS predictions of ^{137}Cs activity concentration in dairy milk and cattle beef are generally underestimated, MRE is 0.47 and 0.60 (Table 5.5), due to the feeding regime implemented in the system. On the other hand, the model seems to predict fairly accurately the radiocaesium concentration in pork and in cereals for the four

years considered, MRE 0.49 and 0.40 (Table 5.5). However, the decrease rate of cereals contamination is lower than the one estimated for the measurements, this might result in an overestimation in the long-term.

5.4.2 Semi-natural

The SAVE prediction of ^{137}Cs activity concentration in roe deer is the most accurate among the models tested, MRE is 1.84, while RIFE1 is characterised by a MRE of 3.63 and the FORM prediction shows a trend which is the opposite of the measurements (Table 5.6; Figure 5.24).

The contamination predicted by SAVE is higher than the observations, for the whole period 1987 – 1991, MRE 1.84. This overestimation is due to the model design, as the radiocaesium concentration in game is quantified using an aggregated transfer factor (T_{agg}, m^2/kg). Therefore the roe deer contamination is estimated as function of the total soil deposition (Bq/m^2). The predicted radiocaesium concentration time dependency is comparable to the observations, half-time 3.7 and 3.9 respectively. Consequently the overestimation is due to the use of a high T_{agg} value.

RIFE1 and FORM are characterised by a similar model design, however the latter estimates the activity concentration in roe deer as function of only the radiocaesium present in the organic and mineral soil layer, while RIFE1 takes also into account the contamination of litter layer. This explains the difference in the initial roe deer contamination for the two models. FORM predicts an initial contamination which is lower than the observations, relative error 0.55, as the ^{137}Cs (Bq/m^2) present in the organic and mineral layer is still low. On the other hand RIFE1 shows a higher contamination, relative error 1.22, since the litter layer is directly exposed to initial deposition.

The two models estimate the roe deer contamination as function of the radiocaesium present in the soil profile, this approach might account for the very slow rate of decline of roe deer contamination predicted by RIFE1 and the build-up estimated by FORM.

The time variation of radiocaesium concentration in game is related to the variation in the total ^{137}Cs available in the soil profile. Nevertheless, ^{137}Cs is very

persistent in soil, (Zibold et al, 2001; Lindner et al, 1994). Therefore, the time variation of radiocaesium activity concentration in soil is very low.

The different prediction trends reported for RIFE1 and FORM are due to the different values of transfer parameters between forest compartments implemented in the two models.

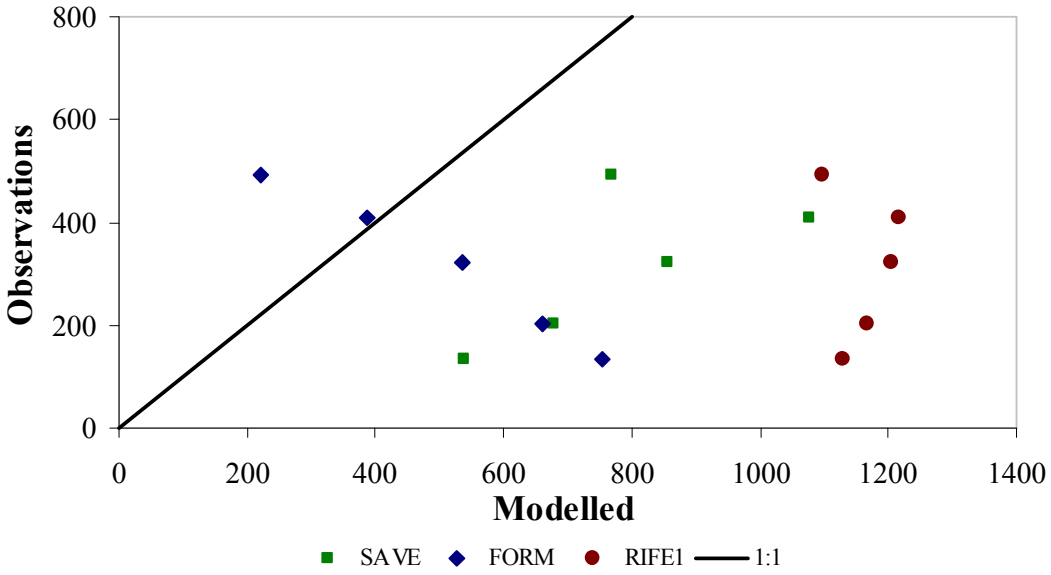


Figure 5.31 The SAVE semi-natural (■), TEMAS semi-natural (FORM) (♦) and RIFE1 (●) model prediction for ^{137}Cs activity concentration in roe deer meat are compared with observations. The solid line represents the 1 to 1 relationship.

5.5 CONCLUSIONS

The comparison between model predictions and the data from the two independent scenarios has been used to determine the level of accuracy of the considered models.

In order to test rural models, i.e. SAVE and TEMAS rural, four years measurements (1986 – 1990) of radiocaesium activity concentration in dairy milk (Bq/l), cattle beef (Bq/kg), pork (Bq/kg) and cereals (Bq/kg) from south Finland have been considered. The semi-natural models, i.e. SAVE semi-nat, FORM (TEMAS) and RIFE1, have been tested evaluating the degree of agreement between the predictions and measurements of ^{137}Cs concentration in roe deer. Samples are from a forest north of lake Constance, Bad Waldsee.

- SAVE provides an accurate prediction from dairy milk and cattle beef, the model has been able to reproduce the seasonal variation in the activity concentration due to changes in feeding regime. However, the prediction of radiocaesium activity concentration in pork and cereals is overestimated, due to the implemented pork diet and to the inaccurate land cover map which is essential to locate the areas for cereal cultivations.

  Regarding the semi-natural test, SAVE provides the most accurate prediction among the models considered, however the radiocaesium concentration in roe deer is overestimated for the considered time period, which is due to the T_{agg} value implemented.

- TEMAS poorly predicts the contamination of dairy milk and cattle beef, nevertheless the degree of agreement between measurements and prediction for pork and cereals is good for the four years considered. In spite of this, an overestimation of cereals contamination might occur in a long-term prediction.

- The estimation of radiocaesium concentration in game provided by FORM is characterised by a trend which is opposite to that shown by the measurements. The model suggested a gradual built-up in the animal body, while the observations illustrate a gradual decrease in function of time. This pattern is caused by the model design, which estimates the animal contamination as function of the soil contamination, i.e. litter, organic matter and mineral layer.

- RIFE1 overestimates the roe deer contamination for the period considered. In addition the prediction does not resemble the observed trend as the decrease rate is 21.5 years, while the measurements show a reduction rate of 3.7 years. This is mainly due to the model design, which estimates the roe deer contamination as a function of the total ^{137}Cs activity concentration in the soil profile. Consequently the game contamination is indirectly affected by the implemented half times between soil compartments.

5.6 SUMMARY

Chapter aim: Accuracy test on the models predictions using two independent scenarios.

- South Finland, rural scenario:

- SAVE. The level of agreement between the model predictions and the observations of ^{137}Cs activity concentration in dairy milk and cattle beef is high, MRE is 0.4 and 0.54 respectively. Nevertheless, the model estimation of pork and cereals contamination is inaccurate, MRE 0.74 and 178 respectively. This low accuracy might be due for the former to the complex animal diet, which the model does not implement, and to the accuracy of the inputs and land cover spatial datasets for the latter.
- TEMAS. The model predictions resemble the measurements of radiocaesium activity concentration in the four food products, i.e. dairy milk, cattle beef, pork and cereals, MRE is 0.47, 0.60, 0.49 and 0.40 respectively. However, the beef cattle and cereals rates of contamination decreases are lower than the observations, which might cause an overestimation of these two food products in the long term.

- Bad Waldsee, semi-natural scenario:

- The level of accuracy of the models prediction is rather low; MRA is 1.84, 1.63 and 3.63 for SAVE, FORM (TEMAS semi-natural) and RIFE1 respectively. Nevertheless, the SAVE prediction trend shows the highest resemblance with the measurements of radiocaesium activity concentration in roe deer. RIFE1 is characterised by a slow decrease rate, which causes an overestimation for the whole considered period. Finally FORM predicts a progressive build-up in the animal tissues, which is the opposite trend to the observations. The low MRE is due to error compensation.

6. ANALYSIS OF MODEL DETAILS AND LEVEL OF COMPLEXITY

The level of detail and complexity are two important model characteristics which can have a strong impact on its applicability and accuracy in prediction.

The level of detail is due to the model design and structure. It can affect the model applicability as each model detail can be regarded as a potential output, and the optimisation process, as model detail can be compared to measurements of system processes and used to "tune" the model parameters.

The level of complexity has a direct effect on the model predictive power and generalizability, which is the capability of a model to produce an accurate prediction using a dataset which was not used for the parameterisation.

However there is no accepted definition of either the level of detail or complexity, which hampers the implementation of any methodologies intended to assess them. Moreover, the literature reports examples of studies where these two terms have been used synonymously, which adds confusion (Stockle, 1992; Webster et al, 1984).

The aim of this chapter is to present a novel approach which supports the analysis of model complexity as function of the level of detail. Four models have been considered: SAVE rural, TEMAS rural, RIFE1 and FORM (TEMAS semi-natural model). The SAVE semi-natural model has not been included in this assessment due to its model design.

The chapter is structured as following:

- Introduction, where an overview of the concepts of model detail and complexity is provided. A brief review of published approaches used to account for model complexity and detail is presented.
- Methodology: a comprehensive overview of a new approach proposed in this work is described. In addition, in order to fully illustrate its potential and characteristics, two sets of models, outside the radioecology field, (Johnson and Thornley, 1983; Johnson and Thornley, 1985; Stapleton, 2004) are analysed, and the results discussed.
- The novel methodology is used to investigate and compare the level of detail and complexity of the four radioecological models.

- This chapter concludes firstly with a discussion of the methodology proposed and secondly reviews the detail level and complexity of the radioelcological models investigated.

6.1 INTRODUCTION

6.1.1 Model detail

Generally the level of detail, when referred to mathematical models, can be defined as "the extent to which the considered systems elements and system relationships are included in the model" (Brooks and Tobias, 1996). Therefore the level of detail that characterises a model can be quantified with the number of model elements which directly refers to system elements. The higher the level of detail the larger the number of processes the model incorporates.

For the purpose of this work this general definition has been further developed and a detail element is considered to be a model variable which can be compared with measurable system processes. Therefore model detail can be compared with observations. Nonetheless, this definition imposes some restrictions on these model variables which can be regarded as "detail elements".

The grass growth model presented by Johnson et al (1985) represents a good illustration of the concept of detail elements. The model describes leaves using four growth stages: new (W_1), mature (W_2), old (W_3) and dead (W_4) and for each growth stage the dry weight is estimated (kg dw). Using the original definition of "level of detail", the four variables, i.e. W_1, W_2, W_3 and W_4, should be regarded as model detail elements. However, there is some subjectivity in the categorisation of new, mature, old and dead leaves, which implies that the measurements of dry weight for each growth stage cannot be considered as absolute measurements.

Therefore it is essential that variables that are considered as model detail elements should refer to measurements which do not have any degree of subjectivity. According to this, the Johnson et al (1985) model would be characterised by only one detail element, regarding leaf dry weight, which is the total leaf dry weight, the sum of W_1, W_2, W_3 and W_4.

The main advantage of detailed models is that they are considered more scientifically credible as they reproduce relevant system processes.

Mathematical models can be divided into two broad categories: mechanistic and empirical or statistical models. Commonly it is assumed that mechanistic models tend to be characterised by a high level of detail as they aim to reproduce the processes of the considered system. By contrast empirical models are often regarded as limited in their level of detail, which however is erroneous. For example, a common practice in radioecological modelling is to quantify radionuclide activity concentration in food products or ecosystem compartments, i.e. wood, leaves etc, using aggregated transfer factors (T_{agg}). T_{agg} values are defined as the ratio of the activity concentration in the considered product (Bq kg^{-1} dw) divided by the total deposition in the soil (Bq m^{-2}). Consequently, a model which uses this approach can estimate the activity concentration in a large number of products, provided that T_{agg} values are available, and therefore it would be characterised by a high level of detail. Nonetheless it is not a mechanistic model as the processes which cause radionuclide uptake, are not considered.

6.1.2 Level of Complexity

The term "model complexity" has a wider interpretation than "level of detail", as it has been used to describe several model aspects. It is possible to identify four categories of model complexity: software, behavioural, structural and mathematical complexity.

Software complexity, considers the logical structure of an algorithm, model implementation, time required by the model simulation and number of code lines present. However, this approach to quantify model complexity is partially dependent on the programming language used, therefore when used for comparing the level of complexity of different models it is limited to software using the same language. Even then it is subject to considerable difficulties.

Behavioural complexity can be regarded as the difficulty in interpreting the model results. An increase in the difficulty of interpreting model outputs is regarded as an increase in the model complexity (Brugnach et at, 2003). Furthermore Flood and Carson in their discussion on complexity define the complexity as "anything that we find difficult to understand" (Brooks and Tobias, 1996). However, assessing the

difficulty of interpreting a model requires a degree of personal judgment and may therefore vary according to the observer.

Structural complexity considers the interaction between model elements. A graphical approach (Figure 6.1) has been proposed by Schruben and Yüceasan (1993) and this has been further developed by the same authors in a following work (Schruben and Yüceasan, 1993). The model complexity is considered to be proportional to the "number of different types of units (entities) in the model, the interaction among units and the number and complexity of decisions made by event routines" (Schruben and Yüceasan, 1993)

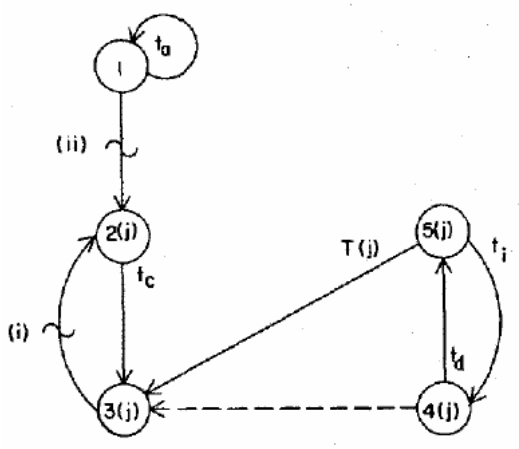


Figure 6.2 Example of an Event Graph referring to a Machine system (Schruben and Yüceasan, 1993).

The main limitation of this methodology is that different event graphs could be produced for the same conceptual model. The authors suggest that the graph having the minimum number of events should be used for the complexity measure. However, there is no certainty that there is only one graph characterised by the minimum number of events. In the case where alternative graphs are present the proposed criterion is not applicable.

Mathematical complexity is based on the mathematical formulation of a model. George (1977) proposed a definition for model complexity which states that "a system can be considered to be complex if it contains sufficient redundancy to be able to function despite the presence of defects", which indicates that a complex model can absorb structural error.

Myung (2000) considers the complexity of a model to be a function of the mathematical formulation adopted, defining model complexity as "the property of a

model that enables it to fit diverse patterns of data" (Pitt and Myung 2002). An intuitive assessment of complexity is the total number of parameters and the mathematical functional form of the relationships implemented, which is a widespread approach to define model complexity in statistics.

Model parameters have been already comprehensively presented in this work (Chapter 3). Generally they are considered as variables in a model equation which can be adjusted, through optimisation, to improve the model fit to the observations. Consider $y = \theta x$, where θ is a model parameter, x is the model input while y is the quantity the model predicts. The mathematical functional form is the way the parameters (θ) and the measurements (x) are combined: the two models $y_1 = \theta x$ and $y_2 = \theta + x$ are characterised by same number of parameters but different functional forms.

The main advantage of quantifying the model complexity as the number of parameters and mathematical formulation is that it creates a linkage between model complexity, model performance, i.e. goodness-of-fit (GOF) predictions-observations, and model generalizability (Myung 2002). In addition, in a complex model the large number of parameters is able to absorb structural error during the optimisation process. Therefore, the quantification of model complexity as number of parameters can be regarded as a different application of the George definition (George, 1977).

The impact that model complexity has on model performance can be illustrated by Figure 6.2a and 6.2b by considering the ranges of data patterns that a model can occupy in the data space.

Data space is defined as the universe of all possible observable data patterns and every point can be regarded as a possible data configuration. Therefore, every model occupies a specific area, where its size is proportional to the model's generalizability.

A simple model (Ms) is characterised by a small region of space (figure 6.2a), as it can describe a limited number of data trends. When those patterns occur the model provides an accurate prediction. Nonetheless, if the considered data are characterised by a pattern which does not belong to the model area, Ms tends to perform poorly.

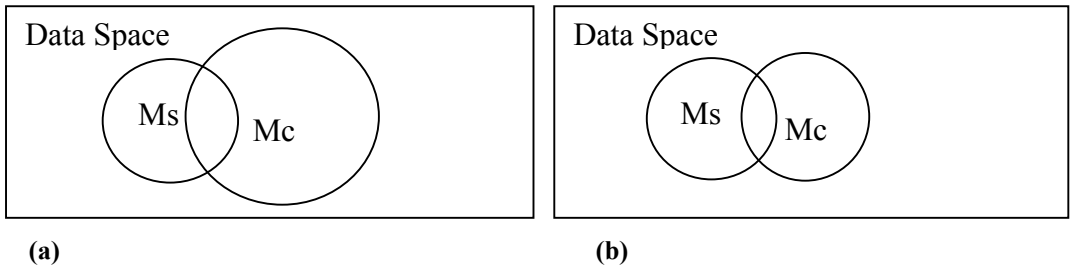


Figure 6.3 Data regions occupied by two models Ms (simple model) and Mc (complex model). In the first example (a) The Mc is characterised by a higher level of generalizability than the Ms due to the larger number of parameters. In contrast in the second example (b), a Mc which is over parameterised might be too specific to the dataset used for the calibration, therefore its level of generalizability is lower and it might perform comparably to a simpler model .

In contrast a complex model (Mc) occupies a larger portion of the data space (Figure 6.2a) as it can describe a larger number of data patterns than Ms. This is mainly due to the number of implemented parameters. The higher the number of parameters, the higher the percentage of data variance the model can explain.

However, a high level of complexity might cause the model to over-fit the considered data. Every data point ($Data_{ob}$) can be considered to be formed by two parts (equation 6.**Error! Reference source not found.**): the real data ($Data_{real}$), which is the holistic measurement of the process investigated and the *error*, which is due to instrumental error, bias in the sampling strategy etc (Brooks and Tobias, 2000).

$$Data_{ob} = Data_{real} + error \tag{6.1}$$

Ideally a model should account only for the real measurements and not be affected by errors, as these are not part of the system investigated and are specific to the dataset considered for the model parameterisation. Therefore, the predictions of a model, characterised by an unjustified high complexity, show a remarkably high GOF when compared to the data used for the optimisation. The consequences of this are reported in Figure 6.3a. This complex model is very likely to describe the combined variance of real data and error and therefore it over-fits the data. This reduces the model flexibility as it is too scenario specific. Consequently, the model is described by a smaller area in Figure 6.2b.

The GOF-complexity and model generalizability-complexity relationships, summarised by Figure 6.4, are not characterised by similar trends. While in the former the increase of GOF is positively related to the complexity, in the latter the generalizability and complexity are positively correlated only up to a particular complexity level, after which the increase of model complexity reduces the model generalizability and the model is considered to be over-parameterised as it over-fits the data.

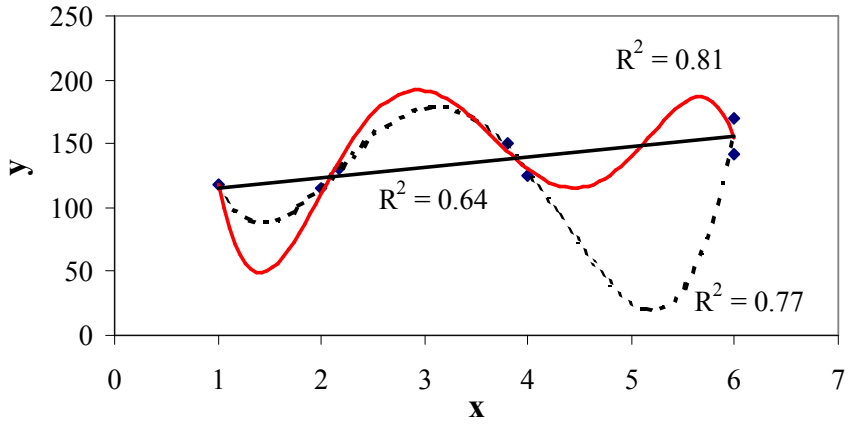


Figure 6.4 The goodness-of-fit of three models (solid black (—), dashed (- - -) and solid red line (—) respectively 2, 4 and 5 parameters. The increase in model fit (R2) is directly proportional to the number of parameters. Although, the most complex model provides the higher level of accuracy it might describe the combined variance of the data and error. This reduces its generalizability as it over-fits the data.

The development of a model characterised by a maximum complexity level, which therefore maximises the model generalizability and GOF, is a delicate and difficult process.

In the field of statistics a common methodology is to produce a family of models and then select a model where the level of complexity is the most justified considering the level of variance of the data. Several indices which can be used to estimate the model generalizability as a function of the model complexity are proposed in the literature: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Minimum Description Length (MIDL), the Information-theoretic measure of complexity (ICOMP) and Cross-Validation (CV). A short description of these is presented below, in order to provide the reader with the state-of-the-art on the subject.

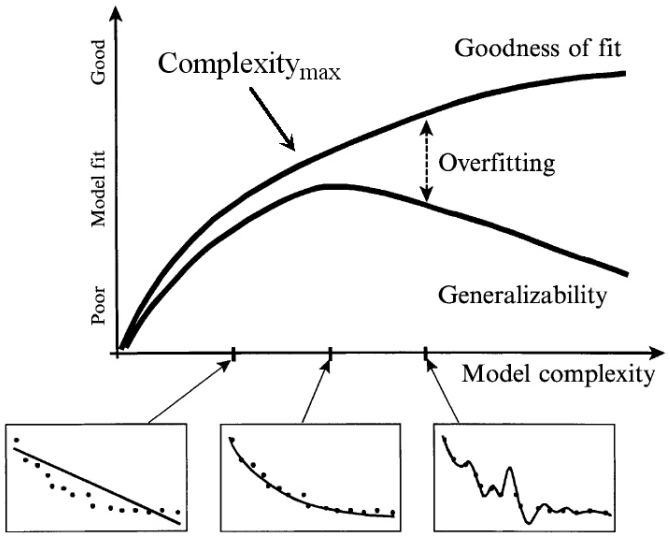


Figure 6.5 The model prediction goodness-of-fit is related to model generalizability as a function of model complexity (Pitt and Myung, 2002).

6.1.2.1 Akaike Information Criterion and Bayesian Information Criterion

The AIC (Bozdogan 2000, Forster 2000; equation 6.9) and BIC (Forster 2000, Zucchini. 2000; equation 6.**Error! Reference source not found.**) index estimate the model generalizability as function of the GOF and the complexity as number of parameters. In addition the BIC includes the sample size.

The first term of the two equations (ln(*ML*); equation 6.9 and 6.**Error! Reference source not found.**) is the natural logarithm of the maximum likelihood (*ML*) of the considered model, therefore it is a measure of goodness-of-fit. The second term refers to the model complexity, p for AIC and $p \cdot \ln(n)$ for BIC, where p is the total number of parameters and n is the number of independent observations that contributes to the likelihood. In both approaches the GOF is compared to the model complexity and the model characterised by the lower AIC or BIC value is selected. A complex model is going to be preferred to a simpler model only if the extra complexity is justified by the model fit.

$$AIC = -2\ln\left(ML\right) + 2p \tag{6.2}$$

$$BIC = -2\ln\left(ML\right) + p\log\left(n\right) \tag{6.3}$$

6.1.2.2 Minimum Description Length and Information-theoretic measure of complexity

The Minimum description length (*MDL*) (Grunwald, 2000) is characterised by two terms: the first is similar to the AIC and BIC index, and it is the natural logarithm of the maximum likelihood function (*ML*). The second term represents a measure of model complexity, which is function of the number of parameters, as in the AIC and BIC but also takes into account the mathematical functional form.

The Information-theoretic measure of complexity (ICOMP) has been developed by Bozdogan (2000) and it is divided into three terms, the first term is a measure of goodness-of-fit, while the second and third terms represent a measure of complexity. For a more detailed overview the Bozdogan (2000) work is suggested.

6.1.2.3 Cross-Validation

The Cross-validation index (*CV*; equation 6.**Error! Reference source not found.**) (Jones, 1994) implements a different approach than the methods previously described. The underlined assumption is that the model generalizability should be based on the model ability to capture the behaviour of unseen observations; therefore the model that produces the most accurate prediction on a blind-validation test should be chosen.

The test uses the residual sum of square as a measure of model goodness-of-fit. The cross-validation procedure consists in randomly removing one observation from the dataset at a time and then comparing the predictions to the excluded measurement (equation 6.**Error! Reference source not found.**).

$$S_{cv}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_{(i)}\right)^2 \tag{6.4}$$

Where $Y_{(i)}$ is the value for the *i*th observation and $\hat{Y}_{(i)}$ is the prediction from the corresponding fit of the model when the *i*th observation is omitted during the model parameterisation.

6.1.2.4 Discussion on the state-of-the-art on model detail and complexity

The overall objective of the indices presented is to select the "best" model among several possible models, however as Zucchini (2000) asserts "an overdose of selection manifests itself in a problem called selection bias". The selection bias refers to a phenomenon which may occur when the selection is made from a large number of comparing models. The main problem is that the random fluctuation of the data, will increase the "score" of some models more than others. Therefore the larger is the family of models the greater is the possibility that the selected model is not the "best" model. In addition, Reid (2004) has observed conflicting responses from the five selection indexes, as they do not select the same model from the model family considered. This experiment introduces a degree of uncertainty on their implementation, arising from the unanswered question of which index is the most appropriate for a given model family.

The state-of-the-art of model detail and complexity could be summarised in the following three points:

- The model detail and level of complexity are two different model characteristics.
- There is not a single accepted definition of model complexity or detail.
- Several indices have been proposed to identify the best model in terms of complexity. However, although desirable, it seems unlikely that a single index can be regarded as the only approach to select a model based on its level of complexity.

Therefore a step forward in the investigation of model complexity could be a structural description of the model which highlights the level of complexity of the main model components. Based on this concept a novel methodology has been developed in this work. It aims to provide an intuitive and user-friendly tool which firstly links model complexity to model details and secondly illustrates the contribution of each detail elements to the overall model complexity.

6.2 METHOD

6.2.1 Detail and complexity matrix

The Detail and Complexity (D&C) matrix (Figure 6.5) is a novel method, developed within the context of this project that aims to be an intuitive and illustrative methodology to create a linkage between model complexity and detail.

The approach adopted to quantify the model complexity is the number of model parameters as for AIC. However not exclusively fitted parameters but also parameters regarded as constant and inputs are considered, unlike AIC. This is due to the difficulty in establishing a clear-cut distinction between truly fixed and adjustable constants.

Model parameters, describing features of the natural world that have been measured to have a variance equivalent to zero, are considered as constant, e.g. charge of an electron. They can be regarded as *truly constant*. However, in many fields of science and especially in environmental science there is a limited number of parameters which could be regarded as truly constant. In most cases, they are derived by some sort of model fitting expert measurement and are characterised by a low level of variance, e.g. k_3 and k_4 in the Absalom model (Chapter 5; Absalom et al, 2001), therefore they can be referred to as *adjustable constants*. Consequently, in the D&C matrix the model complexity is measured as the sum of fitted parameters and constants, due to the difficulty in defining the latter as truly constant.

The number of input parameters has also been included in the D&C matrix as they indirectly affect the model performance and application and therefore is a key model feature. Input parameters introduce uncertainty into the system, which reduces the model prediction accuracy. Nevertheless they are essential to allow the model to provide predictions which are scenario specific.

A D&C matrix (Figure 6.5) is formed by several components: *Detail-Complexity summary*, *Details elements*, *Details characteristics* and *Details connection* elements.

d_t - p_t ,i_t	D_1	D_2	D_3	D_4
D_1	p_1, i_1	x-p_{1-2}, i_{1-2}	p_{1-3}, i_{1-3}	
D_2		P_2, i_2		
D_3			p_3, i_3	
D_4	x-p_{4-1}, i_{4-1}	i_{4-2}		p_4, i_4

Figure 6. 6 Details and Complexity Matrix. The top-left cell reports the number of details elements (d_t), the total number of parameters (p_t) and the number of input (i_t), while in the first column and row the detail elements are described (d_i). In the diagonal the parameters (p_i) and inputs (i_i) required by every detail elements are described. Finally in the off diagonal cells the parameters and inputs shared between two details elements are reported.

The *Detail-Complexity outline* (top left blue cell) reports the values for $\boldsymbol{d_t}$, $\boldsymbol{p_t}$ and $\boldsymbol{i_t}$ which refer respectively to the total number of detail elements, model parameters and inputs.

The *Detail elements* (first column and row, orange cells) report the model details, which are defined as model variables that can be compared to measurable system processes. The order of the detail elements is a crucial aspect of the matrix design, as a result the following hierarchical criteria has been adopted: number of parameters, number of inputs and number of details elements interconnections. However if all these are equal for two or more detail elements then they are arbitrarily ranked, i.e. alphabetically.

The matrix diagonal (red cells) shows the number of parameters and inputs required by each detail element, i.e. $\boldsymbol{p_i}$ and $\boldsymbol{i_i}$ (i = 1 to $\boldsymbol{d_t}$) are the number of necessary parameters and inputs to calculate the details elements $\boldsymbol{D_i}$.

The off diagonal cells (white cells in figure 6.5) are termed *Detail connection* elements and are indicated in the matrix by the $\boldsymbol{x}$ symbol. They have a twofold objective: firstly they indicate the interconnection between details elements. For example, the $\boldsymbol{x}$ present between the $\boldsymbol{D_4}$ and $\boldsymbol{D_1}$ (Figure 6.5) indicates that the detail element $\boldsymbol{D_4}$ is used to calculate the $\boldsymbol{D_1}$ and the same applies to the $\boldsymbol{x}$ between $\boldsymbol{D_1}$ and $\boldsymbol{D_2}$, where $\boldsymbol{D_1}$ is used to in the $\boldsymbol{D_2}$ element. An empty off diagonal cell indicates that the two detail elements are not connected and do not have common parameters or inputs.

Secondly, the detail connection elements show the number of common model parameters ($\boldsymbol{p}$) and inputs ($\boldsymbol{i}$) between two details elements. This information is important to quantify the number of new parameters and inputs that a specific detail

element introduces in the model. For example, detail element D_3 in Figure 6.6 is characterised by two parameters and two inputs (diagonal cell), however, the details connection element $D_2 \frown D_3$ shows that the two parameters have been introduced by D_2, therefore the implementation of D_3 does not increase the overall model complexity.

A D&C matrix provides an outline of the model structure, which supports the user in assessing the complexity in function of the model detail elements. However, it should not be regarded as an equivalent or a substitute to a flow diagram, which represents a complete description of the model structure and interconnection among model components, where a D&C matrix focuses on the model components that can be described as detail elements.

8 - 15, 7	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8
D_1	**5, 5**	2, 3				0, 1		X
D_2	x	**5, 4**	2, 2					
D_3	x	x	**2, 2**			x		
D_4		x	x	**2, 1**				
D_5			x	0, 1	**2, 1**			
D_6						**2, 0**		X
D_7		x	x				**1, 1**	
D_8								**0, 0**

Figure 6.7 D&C matrix of a generic model. The model is characterised by 8 detail elements, 15 parameters and 7 inputs.

6.2.2 D&C matrix application examples

In order, to evaluate the utility of the D&C matrix approach to assess and compare the level of complexity and details of several models, two sets of models will be used: the grass growth models developed by Johnson & Thornley (1983 and 1985) and the terrestrial Carbon-Nitrogen cycle models, "Parent" and "Temp", developed by Stapleton (2004), as they comprise models differing in level of detail and complexity. References to grass '83 and '85 model hereafter will refer to the Johnson & Thornley (1983) and Johnson & Thornley (1985) models respectively.

In the following section a general overview of these models is provided and a more detailed description, i.e. including the mathematical formulation, is reported in Appendix 1, although the original works should be consulted for a full explanation.

6.2.2.1 Grass Growth model

The mechanistic model proposed by Johnson and Thornley (1983) predicts grass growth with inputs of mean daily temperature (°C), daily energy receipt (J m^{-2} d^{-2}) and day length (s d^{-1}).

The main aspect of the grass growth model is the implementation of a new approach to estimate the leaf area expansion. As a result the model can account for the seasonal effects of grazing and cutting treatment on growth and production.

The model can be divided into three sub-models (Figure 6.7):

- *Photosynthesis and growth*, which is used to estimate the rate of synthesis of new structural material above ground as function of photosynthetic process.
- *Above ground structure*, the above ground plant structure is divided into three categories corresponding to growing leaves: first fully expanded leaves, second fully expanded leaves and senescing leaves. This sub-model is implemented to estimate the dry weight (kg carbon m^{-2}) of each of the four leaf categories.
- *Leaf area* establishes the leaf expansion of the four above ground structure categories. The leaf area growth is considered a function of the growth rate.

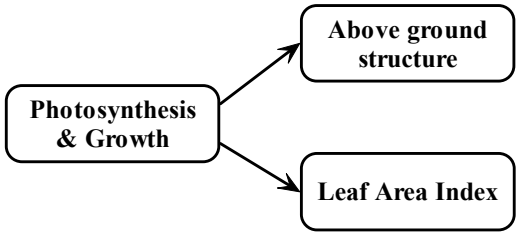


Figure 6.8 The Johnson and Thornley (1983) grass growth model is divided into three sub-models: Photosynthesis and Growth, which is the key components as it is used to estimate the other two sub-models, *Above ground structure* and *Leaf Area Index*

6.2.2.2 Grass Growth model: extended version

The model proposed by Johnson and Thornley (1985) was developed from the model previously described (Johnson and Thornley, 1983). The grass '85 model addresses two important limitations of the previous version, i.e. grass '83 model: the

root system and soil nitrogen, are explicitly incorporated allowing the model to account for the effect of fertilizer application.

The model is summarised in Figure 6.8. The light intercepted by leaves allows photosynthesis to produce the carbon needed for the leaf/shoot and root development. In addition the available nitrogen in soil is considered as a limiting factor for plant growth.

The model can be divided into four sub-models (Figure 6.9):

- *Photosynthesis and growth sub-model*. The photosynthetic process is simulated, the extracted carbon and the available soil N are used to estimate the growth rate of new leaf/shoot and root.
- *Above ground and below ground structure sub-model*. The above and below ground structure, leaf/shoot and root respectively, are divided into four categories representing the transition from growing leaf/shoot and root to senescing. The model estimates the dry weight (kg carbon m^{-2}) of each compartment.
- *Leaf area index sub-model* calculates the leaf area (m^2 m^{-2}) of each of the four age categories based on the growth rate estimated in the Photosynthesis and growth sub-model.
- *Soil Nitrogen sub-model*. The root and leaf/shoot growth rate is considered a function of the nitrogen uptake. Therefore, the soil Nitrogen sub-model estimates the available nitrogen in the soil pool, accounting for the effects of the application of N fertilizer, which is a common feature in grassland farming in the UK (Johnson and Thornley, 1985).

6.2.2.3 C and N cycle in terrestrial ecosystems: the Parent and Temp models

The models described by Stapleton (2004) are an extension of the MBL-MEL model (Rastetter and Shaver, 1997) and are referred to as the Parent and the Temp model (Stapleton, 2004). The models aim to simulate the carbon and nitrogen cycles in terrestrial ecosystems and can be used to estimate the effects that changes in CO_2 concentration, temperature and nitrogen input have on the ecosystem's carbon storage.

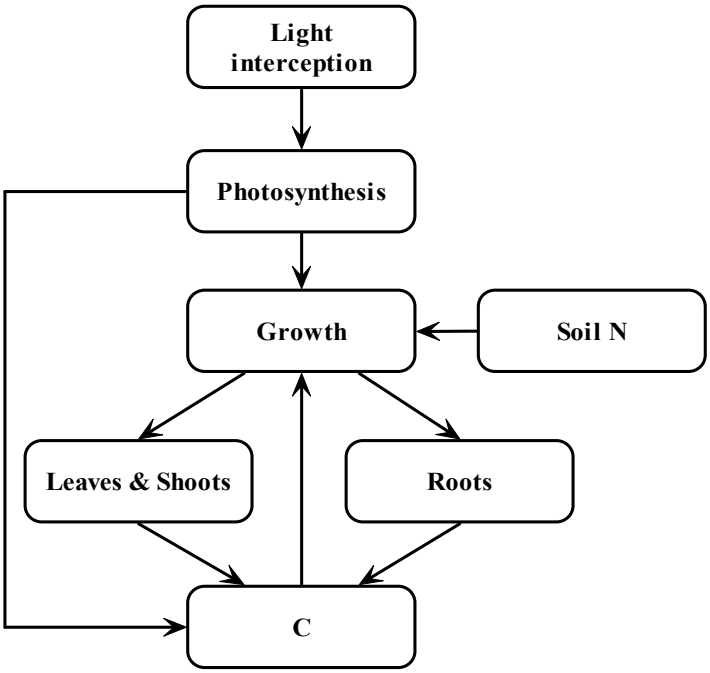


Figure 6.9 The Johnson and Thornley (1985) model introduces two important components: the roots system and the soil nitrogen which limits the plant development and allows the model to take into account the effects of fertiliser application.

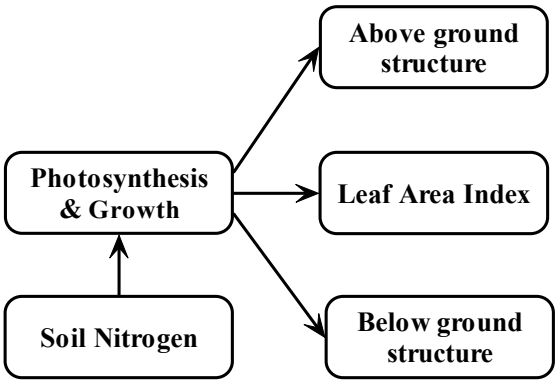


Figure 6.10 The Johnson and Thornley (1985) grass growth model is formed by four sub-models: *Photosynthesis and Growth*, *Above and Below ground structure*, *Leaf Area Index* and *Soil Nitrogen*.

There are three main differences between the MBL-MEL and the Parent model: firstly the former is characterised by a yearly time-step while the latter has been parameterised with data taken once or twice per growing season, which therefore allows the model to have a monthly resolution.

The Parent model can be divided into three sub-models (Figure 6.10):

- *Carbon sub-model*. The carbon absorbed by the vegetation is estimated as function of the CO_2 available and the senescence rate.

- *Nitrogen sub-model*. The estimated inorganic nitrogen available for plant uptake is a function of the plant uptake rate and mineralization/immobilisation due to microbial activity. The microbial biomass is considered to be part of the soil detritus.
- *Acclimation sub-model* allows the model to include the process of acclimation which takes place when the there is a change in the C:N ratio, due to plant development or variations in the ecosystem conditions, i.e. increases of atmospheric CO_2 concentration.

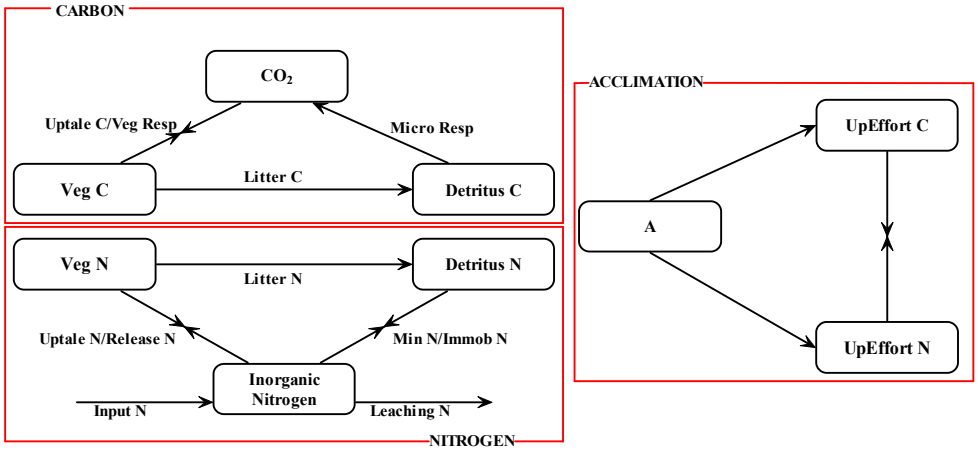


Figure 6.11 Three main sub-models form the Parent model: Carbon, Nitrogen and Acclimation sub-model.

The Temp model, developed by Stapleton (2004), is an extension of the Parent model and it has been designed to account for the temperature dependency of plant physiological processes, which increases the mechanistic credibility of the processes simulated.

6.2.2.4 Discussion

The use of D&C matrices to describe and compare models should allow a user, without any advanced knowledge of the model design and structure, to have an overall understanding on the models differences and to identify the sub-models which account for the highest level of complexity.

The grass growth and C-N fluxes models matrices are reported in Figure 6.11-6.12 and Figure 6.13-6.14 respectively.

Comparing the D&C matrices of the two grass growth models, it is evident that the two models differ substantially in level of complexity and detail. Grass '83 model simulates plants growth considering 6 detail elements with 18 parameters and 12 inputs (Figure 6.11) whereas the grass '85 model has 13 detail elements with 39 parameters and 16 inputs (Figure 6.12).

A comparison of the details elements of the two matrices (Figure 6.11 and 6.12) shows that the grass '85 model includes several elements, such as root system (*Wr* and *Gr*), effect of N and C on the growth rate, which were not included in the grass '83 model. The matrix reflects this model enlargement as total plant dry weight and growth rate is reported as detail elements for both shoots/leafs and roots (W_{sh}, W_r, G_{sh} and G_r).

The grass '85 model D&C matrix shows that the shoots/leafs and roots total dry weight and growth rate (i.e. W_{sh} and W_r, G_{sh} and G_r) are estimated as functions of U_N, which is the nitrogen uptake rate. This implies that the grass '85 model version includes a sub-model to estimate the available nitrogen in soil (N_s). In addition, the matrix shows that N_s is affected by *Na* which is the nitrogen added in the soil during fertilisation, which is a common practice in grassland farming in the UK (Johnson & Thornley, 1985). The nitrogen uptake from the soil nitrogen pool depends on the availability of carbon substrate (*C*). The use of D&C matrix approach is able to highlight this interconnection.

The comparison of the level of complexity of details elements, common to the two models, i.e. leaf area index (*LAI*) and total daily photosynthetic input available for plants (*P*), highlights two important changes in the grass '85 model version.

LAI is estimated in the grass '83 model with 7 parameters and 6 inputs, making it the most complex element of the model and the core sub-model, considering that four of the five details elements are connected to it. In the grass '85 model the *LAI* element shows a lower level of complexity, 5 parameters and 4 inputs. The change in detail elements complexity implies that the approach used to estimate the *LAI* is different for the two models and it has been simplified in the later model.

The grass '83 and '85 models estimate the *LAI* as function of the incremental specific leaf area (η), the rate for leaf appear per tiller (γ) and the rate of production of new shoots/leafs (G_{sh}). However, the grass '83 model includes the W_s (storage dry weight) and the W_g (total structural dry weight), in the calculation of the leaf area index function of the available structure material, while the grass '85 model

implements a more efficient approach. It relates the LAI to the substrate carbon concentration (*C*) and the rate of roots production (G_r), which allows the model to predict increases in *LAI* function of the roots system and indirectly of nitrogen uptake.

Photosynthesis is a primary process in plant growth as it supplies the required energy and carbon. Therefore, the implementation of a more complex sub-model in the grass '85 version, which uses 9 parameters and 3 inputs (Johnson & Thornley, 1984) in comparison with the 6 parameters and 3 inputs of the grass '83 model, is an essential feature of the later model which enables it to account for the fluctuation in daily light and temperature.

The D&C matrices describing the terrestrial Carbon-Nitrogen flux models are reported in figure 6.13 and 6.14. The matrices show that the two models, i.e. Parent and Temp model, do not differ in their level of detail: both are characterised by 15 detail elements. However, there is an increase in the overall level of complexity in the Temp model, which uses 37 parameters and 11 inputs while the Parent model used 19 parameters and 9 inputs. The increase in parameters is mainly related to seven detail elements, i.e. vegetation carbon (*VegC*), Real growth on Carbon (*RG-C*), Nitrogen release (*Nr*), Vegetation uptake of Nitrogen (*VegU-N*), Microbial respiration (*MR*), microbial uptake of Nitrogen (*MU-N*) and gross nitrogen mineralization (*GM-N*). These were considered independent of temperature in the Parent model.

6.2.2.5 Summary

The D&C matrix approach has been found to be effective in providing a methodology to qualitatively compare model detail and complexity, and in presenting an essential overview on the detail elements which affect the model complexity. The use of this approach allows the user to recognize that:

- The grass '83 and '85 models differ substantially in level of detail and complexity.
- The grass '85 model is more complex and is also characterised by a higher level of detail than the grass '83 model. These differences are mainly due to the introduction of new sub-models, as root system and soil nitrogen pool, which increases the model scientific credibility.

- The Parent and Temp model differ substantially in level of complexity although the detail elements are the same.
- The increase in the model complexity between the parent and Temp model is due to the introduction of the temperature dependency for seven processes.

6.3 APPLICATION TO RADIOECOLOGICAL MODELS: RESULTS AND DISCUSSION

6.3.1 Rural models

The detail and complexity matrices for the Absalom model (SAVE DSS; Absalom et al, 2001), TEMAS rural (TEMAS DSS; Montero et al, 2001) and the revisited Absalom model, developed within this project, are reported in Figures 6.14, 6.15 and 6.16, respectively.

The Absalom model has a higher level of detail than TEMAS rural, 9 and 5 detail elements respectively. An examination of the two sets of detail elements reveals that the two models estimate the radiocaesium activity concentration in plants as function of two different soil aspects. The Absalom model estimates the Cs activity concentration in plants using soil chemical characteristics, while TEMAS takes into account the soil physical properties.

The approaches used for the model development, semi-mechanistic and empirical, strongly affect the level of complexity of the two models. The Absalom model is a rather complex model which implements 15 parameters while the TEMAS rural is simpler as it uses only 4 parameters.

The level of complexity can affect the model flexibility, (Chapter 6.1.3), there is an optimum level of complexity for which a model is characterised by the highest degree of flexibility. It is not the aim of this work to assess if the models considered are characterised by the optimum complexity level, nevertheless the low complexity level shown by TEMAS rural model may not be enough to describe the known variability of soil-to-plant transfer of radiocaesium.

6 – 18, 12	*LAI*	*Rm*	*P*	W_{sh}	S_{sh}	G_{sh}
LAI	**7, 6**	3, 1	x-1, 1	4, 1	3, 1	1, 1
Rm	x	**7,5**		x		
P	x		**6, 3**	x		
W_{sh}	x			**4, 6**		x
S_{sh}				x	**3, 5**	
G_{sh}	x	x		x	x	**2, 1**

Figure 6.12 D&C matrix: Johnson & Thornley (1983) growth model. *LAI* is the leaf area index, *Rm* is the maintenance respiration, *P* is the total daily photosynthetic input available for shoot growth, W_{sh} is the total shoots/leafs dry weight, S_{sh} is the daily senescence and G_{sh} is the rate of production of new structure.

13 - 39, 16	*WrT*	*WshT*	*C*	*P*	*Rm*	*Un*	*N*	*Gr*	*Gsh*	*LAI*	*Ns*	*Ssh*	*Na*
Wr	**12, 11**	12, 11	10,10	1, 1	4, 9	2, 5	7, 10	3, 9	3, 9	2, 1	1, 1	2, 5	0, 1
Wsh		**12, 11**											
C			**10, 10**			x		x	x	x			
P	x	x	x	**9, 3**									
Rm	x	x	x		**8, 9**								
Un	x	x	x			**8, 5**	x				x		
N	x	x	x			x	**7, 10**	x	x				
Gr	x	x	x		x	x	x	**5, 9**	x-2, 0	x		x	
Gsh	x	x	x		x		x		**5, 9**				
LAI										**5, 4**			
Ns						x					**5, 1**		2, 0
Ssh												**2, 5**	
Na											x		**2, 1**

Figure 6.13 D&C matrix: Johnson & Thornley (1995) growth model. W_r and W_{sh} are the total roots and shoots/leafs dry weight respectively, *C* is the substrate carbon concentration, *P* is the total daily photosynthetic input available for shoots/leafs and roots growth, *Rm* is the maintenance respiration, U_N is the nitrogen uptake rate, *N* is the is the substrate nitrogen concentration, G_r and G_{sh} are the rate of production of roots and shoots/leafs respectively, *LAI* is the leaf area index, *Ns* is the total soil nitrogen, S_{sh} is the daily senescence and *Na* is the nitrogen available following fertilisation

15 - 19, 9	*VegC*	*RGr-C*	*VegU-N*	*MU-N*	*GMin-N*	*MR*	*Nr*	*Leach-N*	*LitterC*	*LitterN*	*N*	*DetN-*	*Det-C*	*RG-N*	*VegN*
VegC	**7, 4**	x-5,3	x-4,0						x						
RG-C	x	**6, 3**	x												
VegU-N			**6, 0**								x			x	X
MU-N				**4, 0**	x-3,0	x-3,0					x	x			
GMin-N					**3, 0**						x	x			
MR						**3, 0**							x		
Nr							**1, 0**				x			x	X
Leach-N								**1, 0**			x				
LitterC	x								**1, 0**				x		
LitterN										**1, 0**		x		x	X
N			x	x				x			**0, 2**				
Det-N				x	x	x						**0, 1**			
Det-C				x	x	x							**0, 1**		
RG-N	x	x	x											**0, 0**	
VegN	x	x	x				x			x				x	**0, 0**

(a)

15 - 40, 11	*VegC*	*RGr-C*	*VegU-N*	*MU-N*	*GMin-N*	*MR*	*Nr*	*Leach-N*	*LitterC*	*LitterN*	*N*	*DetN-*	*Det-C*	*RG-N*	*VegN*
VegC	**13, 5**	x-11,4	x-4,1			0,1	0, 1		x						
RG-C	X	**12, 4**	x												
VegU-N			**9, 1**								x			x	x
MU-N				**7, 1**	x-3,0	x-3,0					x	x			
GMin-N					**6, 1**						x	x			
MR						**6, 1**							x		
Nr							**4, 1**				x			x	x
Leach-N								**1, 0**			x				
LitterC	X								**1, 0**				x		
LitterN										**1, 0**		x		x	x
N			x	x				x			**0, 2**				
Det-N				x	x	x						**0, 1**			
Det-C				x	x	x							**0, 1**		
RG-N	X	x	x											**0, 0**	
VegN	X	x	x				x			x				x	**0, 0**

(b)

Figure 6.14 D&C matrix of the Parent model (a) and Temp model (b) (Stapleton et al, 2004). *VegC* refers to the vegetation carbon, *RG-C* is the real growth on Carbon, *VegU-N* refers to the vegetation uptake of Nitrogen, *MU-N* is the microbial uptake of Nitrogen, *GMin-N* is the relative N growth rate, *MR* is the microbial respiration, *Nr* refers to the nitrogen release, *Leach-N* is the inorganic Nitrogen depleted through leaching, *LitterC* and *LitterN* is the carbon and nitrogen lost by the plant through senescence, *N* is the inorganic nitrogen in soil, *Det-N* and *Det-C* refer to the nitrogen and carbon located in the top part of the soil profile, *RG-N* the real growth on nitrogen and finally VegN refers to the vegetation nitrogen.

An indication of the limited flexibility of this model is observable in the model-scenario comparison test, Chapter 5. The model predictions for the ^{137}Cs activity concentration in dairy milk, beef, pork and cereals were described by equivalent trends, i.e. high in the first year followed by an approximate equilibrium in the subsequent years. This is not entirely correct as dairy milk, beef and pork contamination are described by a different trend, i.e. a progressive increase is present in the first year, followed by a decline in the second year and subsequently a decline is observable. The low level of complexity reported for TEMAS rural might partially affect the flexibility of the model which is illustrated by its invariable predictions trend.

In contrast the Absalom model shows an adequate degree of flexibility. The predictions resemble closely the observed trends. In addition, for dairy milk and beef measurements on a monthly time resolution the model has been able to account for the periodic increases and decreases which occur within the first and second year.

The overall level of complexity of the revised Absalom model, developed within this project, is higher than the original model (Absalom et al, 2001), 18 and 15 parameters respectively. This increase in the model complexity is due to two factors: the concentration factor (CF) elements which use 4 parameters while the original model 2 and the radiocaesium interception potential (RIP) element is described by 3 parameters. In addition, the implementation of the RIP sub-model allows the revised model to have a higher level of details than the original model, 9 and 8 details elements respectively.

A higher model complexity might be regarded as a potential over-parameterisation of the model which therefore might cause a decrease in the model generalizability. The degree of agreement between the revised model prediction and measurements for soil-to-plant transfer factor (TF), i.e. Smolders and Sanchez dataset, shows an increase, i.e. r^2 0.79 and 0.86 for the original and revised model respectively. Nevertheless, an investigation to verify and quantify the potential reduction in the model flexibility could not be performed due to insufficient independent data. On the other hand, it is important to emphasise that the increase in complexity of the revised model allows a significant reduction in the prediction uncertainty and a more realistic sensitivity to model inputs.

8 - 15,7	$^{137}CsSol$	*kd*	m_k	CEC_h	m_{camg}	*CF*	CEC_c	$^{137}CsPlant$
$^{137}CsSol$	**5,5**	2,3				0,1		x
kd	x	**5,4**	2,2					
m_k	x	x	**2,2**			x		
CEC_h		x	x	**2,1**				
m_{camg}			x	0,1	**2,1**			
CF						**2,0**		x
CEC_c		x	x				**1,1**	
$^{137}CsPlant$								**0,0**

Figure 6.15 D&C matrix of the ^{137}Cs-Plant model developed by Absalom et al (2001). $^{137}CsSol$ is the radiocaesium activity concentration in soil solution, *kd* is the soil kd, m_k is the potassium in soil solution, CEC_h is the Cation Exchange Capacity of the humus soil fraction, m_{camg} is the calcium and magnesium concentration in soil solution, *CF* is the soil-to-plant concentration factor, CEC_c is the Cation Exchange Capacity of the clay soil fraction and $^{137}CsPlant$ is the radiocaesium activity concentration in plants.

5 - 4,21	*Field Capacity*	$CsPlant_{ed}$	$CsPlant_{ined}$	*Cs soil*	*Soil Density*
Field Capacity	**4,3**			x	0,1
$CsPlant_{ed}$		**0,6**	0,4	0,1	
$CsPlant_{ined}$			**0,6**		
Cs soil		x	x	**0,6**	
Soil Density		x	x	x	**0,6**

Figure 6.16 D&C matrix: TEMAS rural model. *Field Capacity* is the soil field capacity, $CsPlant_{ed}$ and $CsPlant_{ined}$ is the radiocaesium activity concentration in the plant edible and inedible part respectively, *Cs soil* is the Cs activity concentration in the soil and *Soil Density* is the soil density.

9 - 18,7	$^{137}CsSol$	CF	kd	RIP	m_k	CEC_h	m_{camg}	CEC_c	$^{137}CsPlant$
$^{137}CsSol$	**4,3**								x
CF		**4,0**							x
kd	x		**3,3**						
RIP	x		x	**3,2**					
m_k	x	x	x		**2,2**				
CEC_h			x		x	**2,1**			
m_{camg}					x		**2,1**		
CEC_c			x		x			**1,1**	
$^{137}CsPlant$									**0,0**

Figure 6.17 D&C matrix of the revisited Absalom model (Absalom et al, 2001) developed within the frame of the present work. $^{137}CsSol$ is the radiocaesium activity concentration in soil solution, CF is the soil-to-plant concentration factor, kd is the soil kd, RIP is the radiocaesium interception potential, m_k is the potassium in soil solution, CEC_h is the Cation Exchange Capacity of the humus soil fraction, m_{camg} is the calcium and magnesium concentration in soil solution, CEC_c is the Cation Exchange Capacity of the clay soil fraction and $^{137}CsPlant$ is the radiocaesium activity concentration in plants.

6.3.2 Semi-natural models

RIFE1 and FORM (TEMAS semi-natural model) show comparable model design, both are dynamic compartment models and the soil is described using three compartments, which refer to litter, organic and mineral layer.

The comparison of the two D&C matrices (Figure 6.17 and 6.18) highlights the higher level of detail and complexity of RIFE1, with 10 and 8 detail elements and 23 and 9 parameters for FORM and RIFE1 respectively. RIFE1 has two additional detail elements than FORM as it includes the tree biomass (kg y^{-1}) and the estimation of understorey contamination. In addition, the plants contamination, i.e. tree, mushroom and understorey, is estimated as a function of the Cs activity concentration in each soil layer. The fractional contribution of each soil layer estimated, taking into consideration the root distribution, partially affects the higher level of complexity of RIFE1 than FORM, which does not include the root system.

The RIFE1 tree biomass model (chapter 2) is a logistic equation (3 parameters and 1 inputs), which simply simulates the plant growth assuming that the growth rate decreases with age. Nevertheless, the radiocaesium activity concentration in the tree is estimated as a function of the tree biomass (kg y^{-1}) and this is an important feature of this model. In addition, RIFE1 could be further developed implementing a more complex tree biomass sub-model, which may apply a more mechanistic approach and consequently be species specific. Several tree growth models have been developed and are reported in the literature, (e.g. Matthews et al 1996). In contrast to RIFE1, FORM does not estimate the tree contamination function of the biomass.

The understorey contamination is not included in FORM. However, it could be easily extended to estimate the radiocaesium activity concentration in understorey as the approach used in this model to calculate vegetation contamination is rather simple (0 parameters and 1 inputs), due to the use of a transfer factor (TF) approach.

10 - 20, 15	*Litter*	*$Tree_{int}$*	*MS*	*OS*	*Mushroom*	*Understorey*	*TreeBio*	*Bark*	*L/N*	*RoeDeer*
Litter	**7, 5**	6,3	3,0	x-4,1	x	x				x
$Tree_{int}$		**6, 3**						x	x	
MS	X	x	**5, 3**	1,1	x	x				x
OS	X	x	x	**5, 3**	x	x				x
Mushroom					**3, 1**					
Understorey						**3, 1**				
TreeBio		x					**3, 1**			
Bark								**1, 1**		
L/N									**1, 1**	
RoeDeer										**0, 1**

Figure 6.18 D&C matrix: RIFE1 semi-natural model. *Litter* is the radiocaesium activity concentration in the litter soil layer, *$Tree_{int}$* is the Cs activity concentration in the tree internal part, i.e. wood, *MS* and *OS* is the radiocaesium contamination of the mineral and organic soil layer respectively, *Mushroom* and Understorey is the radiocaesium activity concentration in these two food products, *TreeBio* is the tree biomass, Bark is the Cs activity concentration in trees bark, *L/N* is the contamination of the tree leafs or needles and *RoeDeer* is the Cs activity concentration in the roe deer body.

8 - 6,6	*Litter*	*MS*	*OM*	*Bark*	*L/N*	*Wood*	*Mushroom*	*RoeDeer*
Litter	**5,1**	1,0	x-1,0	2,1	1,1			
MS		**3,1**	1,0			x	x	x
OM		x	**3,1**			x	x	x
Bark	x			**3,1**				
L/N	x				**2,1**			
Wood						**0,3**		
Mushroom							**0,1**	
RoeDeer								**0,1**

Figure 6.19 D&C matrix: FORM (TEMAS DSS) semi-natural model. *Litter* refers to the Cs activity concentration in the litter soil layer, *MS* and *OM* refer to the radiocaesium contamination of the mineral and organic soil layer respectively, *Bark* and *L/N* are the Cs activity concentration in the tree bark and leaves or needles respectively, *Wood* is the radiocaesium contamination of the wood and *Mushroom* and *RoeDeer* are the Cs contamination of the mushroom and the roe deer body.

6.4 SUMMARY

The level of complexity and detail are two important model features. The former has a strong impact on the model generalizability, i.e. the flexibility of the model to provide a reliable prediction for datasets different from those used for the model optimisation. The model detail affects the model parameterisation process as detail elements can be compared to measurements of system processes in order to "tune" the model parameters. In addition, detail elements can be regarded as supplementary model outputs and therefore increase model applicability.

The methodology proposed in this work, the D&C matrix, has been shown to be a useful "tool" to visualize model complexity, as it can provide the partial complexity level of each detail element, enabling the model structural design to be related to the model complexity.

The implementation of the D&C matrix method on five radioecological models has highlighted the following results:

- The Absalom model (SAVE) is a complex semi-mechanistic model which uses 15 parameters and is characterised by 8 details elements. In contrast the TEMAS rural model is rather simple requiring 4 parameters, although it has 5 detail elements. The higher level of complexity of the Absalom model may explain the high resemblance between the model prediction for dairy milk and beef on the monthly resolution. On the other hand the low level of complexity shown by TEMAS rural model may limit the model implementation due to a low level of generalizability. Although, there is an overall good agreement between the model predictions and the measurements for dairy milk, beef, cereal and pork, the model predicts a trend which does not always resemble the observations.

- The revised Absalom model, developed within the frame of this work, is higher in complexity than the original model (Table 6.1) due to the CF and RIP sub-models. The increase in model complexity may have affected the model generalizability. However, a comparison of model generalizability between the revised and the original Absalom model has not been performed due to lack of independent data.

- RIFE1 and FORM show a strong similarity due to analogous model design. Nevertheless RIFE1 is higher in level of complexity and detail (Table 6.1) due to the tree biomass sub-model and to the root distribution in soil layers, which is used as a weighting factor in the estimation of the Cs uptake by plants from soil.

Table 6.1 Summary of the number of parameters, details elements and inputs for the five radioecological models considered.

Models	Number of		
	Details elements	Parameters	Inputs
Absalom model (SAVE)	8	15	7
TEMAS rural (TEMAS)	5	4	21
Revised Absalom model	9	18	7
FORM (TEMAS)	10	20	15
RIFE1	8	6	6

7. CONCLUSIONS

The initial objective of the present work was to assess five radioecological models in order to provide a detailed overview of their capabilities and limitations. Nevertheless, the literature does not provide a clear set of guidelines or methodologies to perform a model evaluation, which outlines the differences between constructed and conceptual model. A "multi-aspects" comparison approach, which aims to establish these differences, is proposed and used to evaluate the considered models. These results can be used to establish the "best" model among several possible ones.

The process of choosing the best model is a complicated task which requires a comprehensive investigation of different models in order to obtain a full understanding of model performance.

The literature provides several approaches which can be used for model selection. A common method is to estimate the degree of agreement between predictions and measurements. An extension of this, widely used for statistical models is the use of model selection indexes, which assess the model on the basis of both goodness-of-fit and model complexity. However these approaches are not suitable for addressing the question of "how is the resemblance between the constructed and the conceptual model", since they focus only on one single aspect of the models performance, i.e. goodness-of-fit and complexity. They cannot be applied to establish the degree of agreement between the conceptual and constructed model.

The conceptual model can be regarded as the ideal model as it represents what the modelling process aims to develop. The constructed model is the outcome of the modelling project. However, in most cases it does not fully resemble the conceptual model. The differences between them are due to several reasons. For instance, it can be due to the limited understanding of the processes that govern the system investigated or to the mathematical transposition of the conceptual model relationships. Processes, which have a key role in the conceptual model, may have little effect on the prediction variance of the constructed model.

The level of resemblance between the conceptual and conceptual model can only be established by a comprehensive analysis of several model features, including prediction uncertainty, model sensitivity to parameters variance, predictions-

observation goodness-of-fit between predictions and observations, level of complexity and level of detail.

A potential alternative would be the Bayesian Model Averaging (BMA) which does not select a specific model as the process of model selection would do, but instead uses a weighted average of the predictions from the entire family of models. The weight of each prediction is based on the model performance, consequently the fractional contribution of each model might be different for each considered scenario. Although, testing the BMA was not the aim of this work, it could be an important approach to consider in future work with regards to model selection. For a more detailed overview on the BMA relevant literature is suggested, i.e. Gibson et al, 2004, Gelman et al, 2004 and Sivia 1996.

The methodology proposed in this work is divided into two parts. The first considers pragmatic aspects of the model implementations, quality of inputs and outputs, hardware requirements and quality of the software support materials. These aspects of the model evaluation are strongly related to the specification of the project considered and might significantly vary from project to project. The second part, which is the core aspect of this approach, aims to establish the model performance using five different tests: uncertainty, sensitivity analysis, predictions-observations comparison and finally the quantification of the model complexity and level of detail.

- The uncertainty analysis establishes the confidence interval on the prediction. In general terms a large level of uncertainty reduces the reliability of the prediction.
- The sensitivity analysis determines the fractional contribution of model parameters on the prediction variance. Therefore, it highlights the sub-models which significantly affect the model predictions. The results of this test can provide information on the similarity between the conceptual and constructed model.
- The assessment of the accuracy of the model predictions is an essential test in a model testing exercise, however it only provides significant results if the model outputs are compared with independent observations, which therefore have not been used for model parameterisation.

- The level of detail in a model is mainly due to its structure and design. Model details, which are referred in this work as detail elements, could be regarded as secondary model outputs as they describe measurable processes. A detailed model can be applied more widely due to the number of processes that are predicted. On the other hand detailed models can become complex and therefore unwieldy to develop and apply.
- The level of complexity mostly affects model predictive power and generalizability, which is the capability of a model to produce an accurate prediction using an independent dataset.

These methods have been employed to assess and compare five radioecological models.

An overview of the results is given below:

7.1 RURAL MODELS COMPARISON

The Absalom model represents a good illustration of how the comparison between predictions and observations is not sufficient to determine the suitability of a model for a particular scenario application. The uncertainty analysis performed on the model prediction of the TF soil-plant shows a high level of uncertainty, which is mainly due to the high sensitivity of the model to two fitted parameters. Nonetheless the model performed well in the test of prediction accuracy. The estimation of radiocaesium contamination of dairy cow milk and beef shows a high level of resemblance with the observed trends. However the model seems to poorly predict the Cs activity concentration in cereals and pork. This is due to the implementation of the model into the SAVE software rather than a limitation of the underlying approach.

The results obtained from the uncertainty, sensitivity and prediction-observations analysis suggest that the semi-mechanistic design of the Absalom model is able to describe the soil-to-plant flux of radiocaesium. Nevertheless, the main limitation of this model is that its prediction uncertainty is very sensitive to two fitted parameters, i.e k_2 and k_1. This unbalanced model sensitivity can strongly affect the model applicability, as the model prediction accuracy is mostly due to how well these two parameters have been calibrated.

The TEMAS rural model has been developed using a unique design. A comprehensive database is implemented and used to supply the model with a large portion of the required input parameters. The evident advantage of using this type of approach is the reduction in the number of model inputs the user must provide. The uncertainty analysis on three model outputs (Cs activity concentration in soil, edible and inedible part of the plant, Bq/kg) reveals that the model outputs are characterised by a low level of uncertainty, $NG_{75} > 0.6$. In addition the prediction-observations test shows that the TEMAS estimation of radiocaesium activity concentration in rural food products, i.e. dairy milk, beef, cereals and pork, for the four years observations resembles the measurements.

However, the model is particularly sensitive to the global parameters which are part of the database. This is a major weakness of the system as the accuracy of model predictions is entirely based on the quality of the database. The model may only provide an adequate prediction if the scenario data included in the database corresponds to the characteristics of the scenario under investigation.

A further test on the database flexibility has showed that this approach is not able to effectively describe a wide range of scenarios. Therefore an assessment of the database should be performed before any implementation of TEMAS rural.

The Absalom model (SAVE) is a complex semi-mechanistic model which uses fifteen parameters and it is characterised by eight detail elements. In contrast, the TEMAS rural model is rather simple requiring four parameters, although it is formed by five detail elements. Furthermore the number of input parameters clearly shows the difference in the models approaches, SAVE required seven inputs while TEMAS twenty-one, which however are mostly included in the system database.

The higher level of complexity of the Absalom may explain the resemblance detected between the model predictions of Cs activity concentration in dairy milk and beef and the measurements on a monthly resolution. On the other hand the low level of complexity which characterised TEMAS rural might affect the model generalizability.

7.2 OVERVIEW OF THE REVISED ABSALOM MODEL

The uncertainty and sensitivity analysis performed on the Absalom model suggests that the model has little resemblance with the conceptual model, as the

prediction appears to be insensitive to most of the input parameters. Therefore the two sub-models, which have been identified as dominating the prediction uncertainty, have been modified and the model re-parameterised.

The revised model is characterised by a much lower level of uncertainty, in addition the sensitivity analysis shows that local parameters have a key role in the model uncertainty. Finally, there is a significant agreement between the model sensitivity results and the general understanding of processes affecting Cs soil-to-plant transfer. Therefore, the revised model is characterised by a higher level of resemblance with the conceptual model than with the original Absalom model.

These results demonstrate that the uncertainty and sensitivity analysis are crucial tests tools which provide an overview on the model behaviour, which can be used to assess and improve the resemblance between the constructed and conceptual model.

7.3 SEMI-NATURAL MODELS COMPARISON

RIFE1 can be described as a dynamic compartment model and it has been developed to study the radiocaesium fluxes within a semi-natural environment, i.e. soil to vegetation and from upper to lower soil layers. The uncertainty analysis performed on the model prediction of Cs activity concentration in mushrooms, bark and roe deer shows that the model outputs are characterised by a low level of uncertainty. Moreover the model sensitivity highlights the general similarities between the conceptual and the constructed model. Nevertheless the model predicts a slower reduction in Cs activity concentration in roe deer for the five years, which results in largely overestimating the roe deer contamination for the last two years.

FORM is the model implemented in the TEMAS system to estimate the Cs activity concentration in semi-natural products and it is a dynamic compartments model similar in principle to RIFE1. The three outputs investigated, mushroom (Bq/kg), bark (Bq/kg) and roe deer (Bq/kg) are characterised by a low uncertainty. However, unlike RIFE1, the model is mostly sensitive to the transfer factors (TF) used, therefore the model prediction accuracy may be entirely due to the reliability of the TF value adopted.

The estimation of radiocaesium activity concentration in game has been compared against measurements for a five years period. The model estimates a

gradual built up in the animal tissues, however the observations show a progressive reduction. This result demonstrates that the model design is not able to fully account for radiocaesium dynamics.

The SAVE semi-natural model is characterised by the lowest level of complexity among the models considered as it required only one parameter which is the T_{agg} value. Therefore new detail elements can be introduced in the model without changes in model design.

The uncertainty and sensitivity analysis has not been performed as the model prediction uncertainty is directly related to the uncertainty of the T_{agg} value adopted. On the other hand the model estimation of roe deer contamination is characterised by a good level of accuracy, as the prediction resembles the observation trend.

RIFE1 and FORM show several similarities in their model design, e.g. approach adopted to describe the soil layer structure. Nevertheless the detail and complexity matrix highlights the higher level of detail and complexity that characterises RIFE1 with 10 and 8 detail elements and 23 and 9 parameters for RIFE1 and FORM respectively. The main difference is that RIFE1 implements a tree biomass sub-model and therefore estimates the Cs activity concentration in trees as function of the plant mass. This is not included in FORM. In addition RIFE1 estimates the contamination of vegetation, i.e. tree, mushroom and understorey, as function of the Cs activity concentration in each soil layer. The fractional contribution of each soil layer to the overall plant contamination is weighted taking into consideration the vertical distribution of the root system. Nonetheless, the two models seem to use a similar approach to estimate the Cs activity concentration in roe deer.

The selection of the best model is a difficult task which requires a deep knowledge of the models structure and performance. The estimation of the degree of agreement between the predictions and observations does not provide sufficient information to establish the true model performance. As demonstrated in this work, models that show a high level of prediction accuracy may fail to provide outputs with a low uncertainty or with an adequate sensitivity to input parameters.

The models investigated show a limited resemblance with their conceptual models. This is due to low predictions accuracy, i.e. RIFE1 and FORM, high level of uncertainty (e.g. SAVE rural model) or sensitivity to model parameters which are not

in agreement with the current understanding of radiocaesium behaviour in the environment, (e.g. TEMAS and SAVE rural).

A potential alternative is by contrast not to select a specific model, but to use a weighted average of the predictions from the entire family of models.

The results of the present work have highlighted the difficulties that are present in using methodologies to select the "best model".

The results have led to the conclusion that a reliable methodology for model selection should be based on a comprehensive investigation of each considered model aspect, as there is not a single best approach. The methodology proposed in this work has managed to fully accomplish such a task in the case of the five radioecological models studied.

APPENDICES

A.1 GRASS '83 MODEL

The mechanistic model proposed by Johnson and Thornley (1983) predicts grass growth using mean daily temperature (T; °C), daily energy receipt (J; J m^{-2} d^{-2}) and day length (h; s d^{-1}).

The main aspect of the grass growth model is the implementation of a new approach to estimate the leaf area expansion. As a result the model can account for the seasonal effects of grazing and cutting treatment on growth and production.

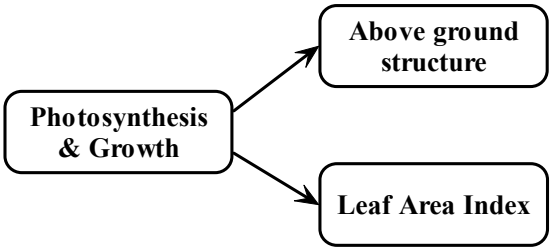


Figure A.20 The Johnson and Thornley (1983) grass growth model is divided into three sub-models.

The model can be divided into three sub-models (Figure A.1):

- *Photosynthesis and growth*, which is used to estimate the rate of synthesis of new structural material above ground as function of photosynthetic process.
- *Above ground structure*, divided into three categories corresponding to growing leaves: first fully expanded leaves, second fully expanded leaves and senescing leaves. This sub-model is implemented to estimate the dry weight (kg Carbon m^{-2}) of each of the four leaves categories.
- *Leaf area index* sub-model establishes the leaves expansion of the four above ground structure categories. The LAI is considered function of the growth rate.

A.1.1 Photosynthesis and growth sub-model

The daily canopy gross photosynthetic rate (P, kg Carbon m^{-2} d^{-1}) is estimated by equation A.1. This approach does not take into consideration the relationship present between plant photosynthetic efficiency to leaf age, demonstrated by the Woledge and Leafe work (Johnson and Thornley, 1983), and to specific leaf area. Nevertheless it accounts for the temperature as a limiting factor, the light-saturated leaf gross photosynthetic rate (P_m, kg CO_2 m^{-2} s^{-1}; equation A.2) is positively correlated to the mean daily temperature (T).

The main input to estimate the plant gross photosynthetic rate is the daily energy receipt (J m^{-2} d^{-1} PAR) by the plant (J) as solar radiation. In addition the energy attenuation due to the canopy density is accounted, τ and k are two parameters which indicate leaf transmission for the former and extinction coefficient for the latter.

$$P = \frac{P_m h}{k}\left[\frac{\alpha k J/h + (1-\tau)P_m}{\alpha k J^{-kL/h} + (1-\tau)P_m}\right]\theta \qquad \text{(A.1)}$$

where,

α is the leaf photosynthetic efficiency, $12*10^{-9}$ (CO_2 J^{-1})

k is the extinction coefficient of canopy, 0.5

τ is the Leaf transmission coefficient, 0.1

h is the daylength (s d^{-1})

J is the daily energy (PAR) receipt (J m^{-2} d^{-1})

L is the cumulative leaf area index (m^2 m^{-2})

θ is a factor to convert the CO_2 to carbon; taken to be the ratio of the relative molecular masses of carbon and CO_2 12/44 (kg Carbon kg CO_2^{-1})

$$P_m = P_{20}\frac{T - T_0}{20 - T_0} \qquad \text{(A.2)}$$

where,

T is the mean daily temperature (°C)

T_0 is the temperature at which crops growth stops, 0 (°C)

P_{20} is the light-saturated gross photosynthetic rate at 20 °C, $1.5*10^{-6}$ (kg CO_2 m^{-2} s^{-1})

The plant uses the carbon fixed, by the photosynthetic process, for three purposes: shoot and roots growth and for maintenance. Therefore, Φ, ranging between 0 and 1, is the fraction of plant photosynthesis allocated to the shoots growth.

The rate of synthesis of new structural material (G) is based on the model proposed by Thornley and Hurd (Johnson and Thornley, 1983) (equation A.3; kg C m^{-2} d^{-1}). The actual plant growth is positively correlated to the plant storage dry weight (W_s, kg carbon m^{-2}; equation A.10) and to the total structure dry weight (W_G, kg carbon m^{-2}; equation A.11), however the total crop weight (W), which is the sum of W_G and W_s, is considered as a limiting factor, therefore it is negatively correlated to G (equation A.3).

The mean daily temperature can also inhibit the plant development; μ is the growth coefficient (d^{-1}; equation A.4) which accounts for the temperature effect.

$$G = \mu \frac{W_s W_G}{W} \tag{A.3}$$

$$\mu = \mu_{20} \frac{T - T_0}{20 - T_0} \tag{A.4}$$

where,

μ_{20} is the growth coefficient at 20°C, 0.5 d^{-1}

T is the mean daily temperature (°C)

T_0 is the temperature at which crops growth stops, 0 (°C)

A yield factor (Y) is defined as the units of structure that results from synthesis as one unit of storage material, the rest being respired. A typical value for Y is 0.75. Consequently, the rate of utilization of storage for the production of new structure is represented by the ratio between G and Y.

The maintenance respiration has been observed to be proportional to the plant dry weight, although the proportionality is reduced with plant age. Consequently, in the Johnson and Thornley (1983) grass growth model, it is assumed that the maintenance costs per dry weight is different for each of the four compartments, representing the four leaf growth stages. Equation A.5 estimates the maintenance respiration (R_m) function of the maintenance respiration coefficient (M_i) and the components of structural dry weight (W_i), where i ranging between 1 and 4 represents the four growing stages.

$$R_m = \sum_{i=1}^{4} M_i W_i \qquad \text{(A.5)}$$

A.1.2 Leaf Area index sub-model

The growth G, estimated in equation A.3 is the total plant growth, nevertheless a fraction of it, i.e. ρ ranging between 0 and 1, is used for leaves growth while the (1 - ρ) fraction is used for stems development.

Johnson and Thornley (1983) assume a constant ρ equal to 0.7, based on the Robson (Johnson and Thornley, 1983) measurements, which highlight that the ratio between leaf and stem weight is almost constant for the entire growing period.

The rate of production of new leaf area is estimated as the product between the incremental specific leaf area (δ, m^2 of leaf kg $Carbon^{-1}$) and ρG. Where δ is estimated by the equation A.6. Johnson and Thornley regard this approach as consistent with observations, although the authors admit there "there aren't sufficient data available to fully justify the specific form of the equation adopted".

$$\delta = \delta_m \left(1 - \frac{W_s}{W} \right) \qquad \text{(A.6)}$$

where,

δ_m is the maximum value for δ, 40 m^2 leaf kg $Carbon^{-1}$

The leaf area index of each of the four compartments, which represent the growing leaves, first fully expanded leaves, second fully expanded leaves and

senescing leaves, is estimated by the set of differential equations reported previously (equation A.7a – A.7d).

$$\frac{dL_1}{dt} = \delta\rho G - \lambda\gamma L_1 \quad \text{(A.7a)}$$

$$\frac{dL_2}{dt} = \lambda\gamma L_1 - \gamma L_2 \quad \text{(A.7b)}$$

$$\frac{dL_3}{dt} = \gamma L_2 - \gamma L_3 \quad \text{(A. 7c)}$$

$$\frac{dL_4}{dt} = \gamma L_3 - \gamma L_4 \quad \text{(A. 7d)}$$

where,

λ is a weighting factor for the flux of material from first to second compartment.

γ is rate of leaf appearance per tiller, d^{-1} (equation A.8)

$$\gamma = \gamma_{20} \frac{T - T_0}{20 - T_0} \quad \text{(A.8)}$$

where,

T is the mean daily temperature (°C)

T_0 is the temperature at which crops growth stops, 0 (°C)

γ_{20} is the rate of leaf appearance at 20°C, 0.15 (d^{-1})

A.1.3 Above ground structure sub-model

The above ground structure sub-model is formed by 5 main components: W_i (i=1– 4) (equation A.9a – A.9d) and W_s (equation A.10), which estimate the dry weight (kg Carbon m^{-2}) of the four leaf compartments and the dry weight of the carbon storage compartment (kg Carbon m^{-2}), respectively.

$$\frac{dW_1}{dt} = G - \lambda\gamma W_1 \quad \text{(A.9a)}$$

$$\frac{dW_2}{dt} = \lambda\gamma W_1 - \gamma W_2 \quad \text{(A.9b)}$$

$$\frac{dW_3}{dt} = \gamma W_2 - \gamma W_3 \quad \text{(A.9c)}$$

$$\frac{dW_4}{dt} = \gamma W_3 - \gamma W_4 \quad \text{(A.9d)}$$

where,

γ is rate of leaf appearance per tiller, d^{-1} (equation A.8)

$$\frac{dW_s}{dt} = P - G/Y - R_m \quad \text{(A.10)}$$

where,

P is the canopy gross photosynthetic rate (kg Carbon m^{-2} d^{-1}; equation A.1)

G is the rate of production of new structure, i.e. leaf and stem (kg Carbon m^{-2} d^{-1}; equation A.3)

Y is the crop yield factor, 0.75

R_m is the maintenance respiration (kg Carbon m^{-2} d^{-1}; equation A.5)

$$W_G = W_1 + W_2 + W_3 + W_4 \quad \text{(A.11)}$$

A.2 GRASS '85 MODEL

The grass growth model proposed by Johnson and Thornley (1985) was developed from the model previously described (Johnson and Thornley, 1983). The new model addresses two important limitations of the previous version: roots system

and nitrogen in soil, which allow the model to account for the effect of fertilizer application.

The model can be summarised by figure A.8, where the light intercepted by leaves allows the photosynthetic process to produce the carbon needed for the leaf/shoot and root development. In addition the available Nitrogen in soil is considered a limiting factor for plant growth.

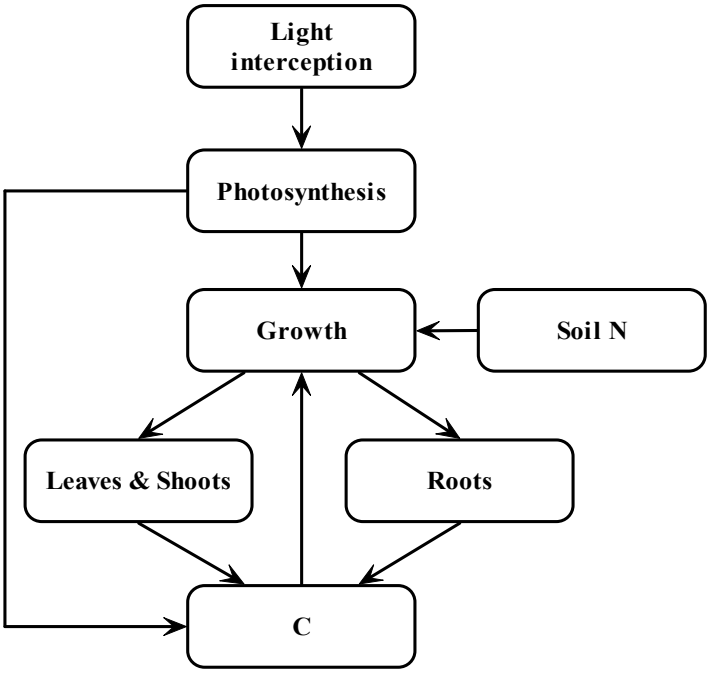


Figure A.21 The Johnson and Thornley (1985) model introduces two important components: the roots system and the soil nitrogen which limit the plant development and allow the model to take into account the effects of application of fertiliser.

The model can be divided into five sub-models (Figure A.3):

- *Photosynthesis and growth sub-model*. The photosynthetic process is simulated, the extracted Carbon and the available soil N are used to estimate the growth rate of new leaf/shoot and root.
- *Above ground and Below ground structure sub-model*. The above and below ground structure, leaf/shoot and root respectively, are divided into four categories representing the transition from growing leaf/shoot and root to senescing. The model estimates the dry weight (kg Carbon m^{-2}) of each compartment.

- *Leaf area index sub-model* calculates the leaf area (m^2 m^{-2}) of each of the four age categories based on the growth rate estimated in the Photosynthesis and growth sub-model.
- *Soil Nitrogen sub-model*. The roots and leaf/shoot growth rate is considered function of the nitrogen uptake. Therefore, the soil Nitrogen sub-model estimates the available nitrogen in the soil pool and consequently it accounts for the effects of the application of N fertilizer, which is a common feature in grassland farming in the UK (Johnson and Thornley, 1985).

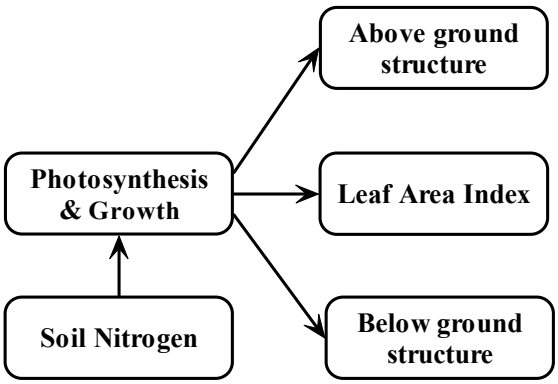


Figure A.22 The Johnson and Thornley (1985) grass growth model is formed by four sub-models: Photosynthesis and Growth, Above and Below ground structure, Leaf Area Index and Soil Nitrogen.

A.2.1 Photosynthesis and growth sub-model

The sub-model used to estimate the daily canopy photosynthesis in the Johnson and Thornley (1983) model has been replaced with a more complex canopy photosynthesis model, a more detailed description can be found in the original Johnson and Thornley (1984) publication.

The single leaf gross photosynthetic rate (P_g, kg CO_2 m^{-2} s^{-1}) is given by equation A.12.

$$P_g = \frac{1}{2\theta}\left[\alpha I_l + P_m - \sqrt{\left(\alpha I_l + P_m\right)^2 - 4\theta\alpha I_l P_m}\right] \qquad \text{(A.12)}$$

Where θ is a dimensionless parameter ($0 \leq \theta \leq 1$), α is the photosynthetic efficiency (kg CO_2 J^{-1}), I_l is the light flux density on the leaf (J m^{-2} s^{-1}) and P_m is the

asymptotic value of P_g at high light (equation A.13). Consequently, the gross photosynthetic rate for the whole canopy is calculated by integrating P_g for the entire canopy (equation A.14).

$$P_m = P_{20} \frac{T - T_0}{20 - T_0} \tag{A.13}$$

where,

T is the mean daily temperature (°C)

T_0 is the temperature at which crops growth stops, 0 (°C)

P_{20} is the light-saturated gross photosynthetic rate at 20 °C, $0.6*10^{-6}$ (kg CO_2 m^{-2} s^{-1})

$$P_c = \int_0^L P_g \left(I_l \right) dL \tag{A.14}$$

where,

L is the leaf area index

I_l instantaneous light flux density incident on the leaves at leaf area index L (equation A.15; W m^{-2})

P_g single leaf gross photosynthetic rate (kg CO_2 m^{-2} s^{-1})

$$I_l = \left(\frac{k}{1 - \tau} \right) I_0 e^{-kL} \tag{A.15}$$

Where k is the canopy extinction coefficient and τ is the leaf transmittance. Therefore, the total daily photosynthetic input (P; kg Carbon m^{-2} d^{-1}) is equal to P_g integrated for the daylengh (h; s d^{-1}; equation A.16). The ratio between the carbon and CO_2 atomic mass, i.e. 12/44, is a factor to convert from CO_2 to carbon.

$$P = \frac{12}{44} \int_0^h P_c dt \tag{A.16}$$

This approach to estimate the total photosynthetic input (equation A.16) does not take into consideration the effect of nitrogen-deficiency, which is known to be a limiting factor particularly in high light levels (Johnson and Thornley, 1985).

Nevertheless, the main effect to growth in nitrogen-deficient plants is the lack of nitrogen, therefore any variations in the photosynthetic parameters can be regarded as unimportant.

The model used to estimate the leaf/shoot and root growth rate is described in more details in Johnson and Thornley,(1985). The rate of synthesis of new leaf/shoot and root structure, G_{sh} and G_r (kg structure m^{-2} d^{-1}) respectively, are function of the substrate carbon and nitrogen concentration and the total leaf/shoot and root structure dry weight (equation A.17 and A.18).

$$G_{sh} = \mu CN \lambda_{sh} W_{sh} \tag{A.17}$$

$$G_r = \mu CN \lambda_r W_r \tag{A.18}$$

Where μ is a growth coefficient (d^{-1} ([C][N])$^{-1}$; equation A.19) and λ_{sh} and λ_r are the fractions that determine the partitioning between new growth of leaf/shoot and root (equation A.20 and A.21)

$$\mu = \mu_{20} \frac{T - T_0}{20 - T_0} \tag{A.19}$$

where,

μ_{20} is the growth coefficient at 20°C, 150 (d^{-1} ([C][N])$^{-1}$)

T is the mean daily temperature (°C)

T_0 is the temperature at which crops growth stops, 0 (°C)

$$\lambda_{sh} = \frac{\xi N f_r}{\xi N f_r + C f_{sh}} \tag{A.20}$$

$$\lambda_r = \frac{C f_{sh}}{\xi N f_r + C f_{sh}} \tag{A.21}$$

Where ξ is the partitioning parameter (50 kg carbon (kg nitrogen)$^{-1}$), C and N are the substrate carbon and nitrogen concentrations, based on the live structure dry weight

(equation A.22 and A.23). f_{sh} and f_r are the leaf/shoot and root live structure fractions (equation A.24 and A.25). Considering that the sum between λ_{sh} and λ_r is 1, G_{sh} and G_r are mainly determined by μ, therefore by the mean daily temperature.

$$C = \frac{W_C}{W_G} \tag{A.22}$$

$$N = \frac{W_N}{W_G} \tag{A.23}$$

$$f_{sh} = \frac{W_{sh}}{W_G} \tag{A.24}$$

$$f_r = \frac{W_r}{W_G} \tag{A.25}$$

where,

W_C,W_N are the substrate Carbon and Nitrogen, kg Carbon m^{-2} and kg Nitrogen m^{-2} respectively.

W_{sh}, W_r are the total live leaf/shoot and root structural dry weight, kg structural dry weight m^{-2}.

W_G is the total live crop structural dry weight, kg structural dry weight m^{-2}.

The shoot/root maintenance respiratory ($R_{m,sh}$) is estimated using a similar approach implemented in the previous model version (Johnson and Thornley, 1983) (equation A.26)

$$R_{m,sh} = f_C \sum_{i=1}^{4} \left(M_{sh,i} W_{sh,i} + M_{r,i} W_{r,i} \right) \tag{A.26}$$

Where, f_C is the fractional carbon contents in live plant structure, 0.45 kg carbon (kg structure)$^{-1}$, M_{sh} and M_r are the maintenance coefficient of leaf/shoot and root structural components (equation A.27 and A.28 respectively)

$$M_{sh,i} = M_{20sh,i} \frac{T - T_0}{20 - T_0} \quad \text{(A.27)}$$

$$M_{r,i} = M_{20r,i} \frac{T - T_0}{20 - T_0} \quad \text{(A.28)}$$

where,

$M_{20sh,i}$ and $M_{20r,i}$ are the Maintenance coefficients of leaf/shoot and root at 20°C, i = 1 – 4: 0.02, 0.02, 0.015 and 0.01 d^{-1}.

T is the mean daily temperature (°C)

T_0 is the temperature at which crops growth stops, 0 (°C)

A.2.2 Above ground and below ground structure sub-model

The above ground plant structure is divided into four compartments: growing leaves (W_{sh1}), first fully expanded leaves (W_{sh2}), second fully expanded leaves (W_{sh3}) and senescing leaves (W_{sh4}). This approach is similar to the one adopted by Johnson and Thornley (1983), and it allows to describe the leaf turnover and senescence. The total leaf/shoot dry weight is estimated as the sum of the dry weigh of each age compartment (equation A.29).

$$\sum_{i=1}^{4} W_{sh,i} \quad \text{(A.29)}$$

The fourth compartment (W_{sh4}) comprises the leaves that are senescing. Nevertheless, assuming that on average half of the leaves included in this categories are dead, the total live leaf/shoot structure dry weight is determined by equation A.30

$$W_{sh} = \sum_{i=1}^{4} W_{sh,i} - \frac{W_{sh4}}{2} \quad \text{(A.30)}$$

The turnover of the below ground plant structure is described in an analogue way to the leaf/shoot, four age compartments representing the transition from growing to senescing roots (W_{r1}, W_{r2}, W_{r3} and W_{r4}) and the total below ground plant structure is calculated by equation A.31, which derived that the total live root structure dry weight is estimated by equation A.32.

$$\sum_{i=1}^{4} W_{r,i} \quad \text{(A.31)}$$

$$W_r = \sum_{i=1}^{4} W_{r,i} - \frac{W_{r4}}{2} \quad \text{(A.32)}$$

The total live crop structure dry weight (kg structural dry weight m^{-2}) is considered as the sum of the total leaf/shoot and root weight (equation A.33).

$$W_G = W_{sh} + W_r \quad \text{(A.33)}$$

The destination between the live and dead structure was not included in the previous model version (Johnson and Thornley, 1983) however this new approach is consistent with several observations reported in the literature (Johnson and Thornley, 1985).

There are two substrate components: carbon (W_C, kg carbon m^{-2}) supplied by the photosynthetic process and nitrogen (W_N, kg carbon m^{-2}) supplied by the roots uptake. The total storage dry weigh is therefore estimated as reported by equation A.34.

$$W_S = \frac{M_C}{12} W_C + \frac{M_N}{14} W_N \quad \text{(A.34)}$$

Where M_C and M_N are the molecular masses of carbon (^{12}C) and nitrogen (^{14}N), while W_C and W_N are estimated by the differential equation A.35 and A.36 respectively.

$$\frac{dW_C}{dt} = P - \frac{f_C}{Y}\left(G_{sh} + G_r\right) - R_m - R_N + D_C \quad \text{(A.35)}$$

$$\frac{dW_N}{dt} = U_N - f_N\left(G_{sh} + G_r\right) + D_N \quad \text{(A.36)}$$

where,

P is the total daily photosynthetic input, kg carbon m^{-2}

f_C and f_N is the fractional carbon and nitrogen content of live plant, 0.45 and 0.025 kg carbon (kg structure)$^{-1}$ respectively

Y yield factor for structural growth (0.75)

G_{sh} and G_r rate of production of new leafs/shoots and roots (kg structural dry weight m^{-2} d^{-1})

R_m rate of maintenance respiration (kg C m^{-2} d^{-1})

R_N respiration associated to N uptake (kg C m^{-2} d^{-1})

D_C and D_N supply of substrate C and N from degraded shoots and roots structure.

The dry weight of each of the four compartments, which describe the leaf/shoot at different stage of development is estimated using the same approach implemented in the previous model version (equation A.37a – A.37d) (Johnson and Thornley, 1983).

$$\frac{dW_{sh1}}{dt} = G_{sh} - 2\gamma_{sh}W_{sh1} \quad \text{(A.37a)}$$

$$\frac{dW_{sh2}}{dt} = 2\gamma_{sh}W_{sh1} - \gamma_{sh}W_{sh2} \quad \text{(A.37b)}$$

$$\frac{dW_{sh3}}{dt} = \gamma_{sh}W_{sh2} - \gamma_{sh}W_{sh3} \quad \text{(A.37c)}$$

$$\frac{dW_{sh4}}{dt} = \gamma_{sh}W_{sh3} - \gamma_{sh}W_{sh4} \quad \text{(A.37d)}$$

The compartments describing the roots development stage are analogue to the leaf/shoot, equation A.38a – A.37d.

$$\frac{dW_{r1}}{dt} = G_r - 2\gamma_r W_{r1} \quad \text{(A.38a)}$$

$$\frac{dW_{r2}}{dt} = 2\gamma_r W_{r1} - \gamma_r W_{r2} \quad \text{(A.38b)}$$

$$\frac{dW_{r3}}{dt} = \gamma_r W_{r2} - \gamma_r W_{r3} \quad \text{(A.38c)}$$

$$\frac{dW_{r4}}{dt} = \gamma_r W_{r3} - \gamma_r W_{r4} \quad \text{(A.38d)}$$

Where γ_{sh} and γ_r are the rate of leaf/shoot and root appearance per tiller and root turnover respectively and they are positively correlated to the mean daily temperature (equation A.39 and A.40)

$$\gamma_{sh} = \gamma_{sh20} \frac{T - T_0}{20 - T_0} \quad \text{(A.39)}$$

$$\gamma_r = \gamma_{r20} \frac{T - T_0}{20 - T_0} \quad \text{(A.40)}$$

where,

γ_{sh20} and γ_{r20} are the rate of leaf/shoot and root appearance at 20°C, 0.15 d^{-1} and 0.03 d^{-1}

T is the mean daily temperature (°C)

T_0 is the temperature at which crops growth stops, 0 (°C)

A.2.3 Soil Nitrogen sub-model

Nitrogen uptake rate by the roots is considered positively correlated to the total live root structure dry weight ($W_{r,i}$, kg structure dry weight m^{-2}) (equation A.41). Nevertheless, the use of equation A.41 and A.42 does not imply that the nitrogen uptake by the plant is directly proportional to the inorganic nitrogen present in the soil pool (N_s) as C and N may be influenced by different levels of N_s (equation A.22 and A.23 respectively).

$$U_N = \frac{\sigma_N \left(W_{r,1} + \nu_2 W_{r,2} + \nu_3 W_{r,3} + \nu_4 W_{r,4} \right)}{1 + \frac{K_C}{C}\left(1 + \frac{N}{K_N} \right)} \tag{A.41}$$

$$\sigma_N = \tilde{\sigma}_N N_S \tag{A.42}$$

$$\tilde{\sigma}_N = \tilde{\sigma}_{20N} \frac{T - T_0}{20 - T_0} \tag{A.43}$$

where,

$\tilde{\sigma}_{20N}$ is the roots activity parameters at 20°C, $3*10^3$ kg dry soil (kg root structural dry weight)$^{-1}$ d^{-1}

T is the mean daily temperature (°C)

T_0 is the temperature at which crops growth stops, 0 (°C)

N_s soil inorganic Nitrogen content, kg nitrogen (kg dry soil)$^{-1}$

$W_{r,i}$ (i = 1 – 4) are the live structural dry weight , k structural dry weight m^{-2} (equation A.38a – A.44d)

v_i (i = 2 – 4) is the specific root activity weighting parameters (dimensionless), 0.5, 0.25 and 0.1

K_C and K_N are the root activity parameters, 0.01 [C], 0.005 [N]

C and N substrate carbon and nitrogen concentration, kg carbon/nitrogen (kg structural dry weight)$^{-1}$

The rate of plant uptake of inorganic soil nitrogen (U_N) is only function of the soil Nitrogen content (N_s, kg nitrogen (kg dry soil)$^{-1}$. This may be regarded as a

limitation of the Nitrogen sub-model as there are several processes which should be taken into consideration to determine the available nitrogen to plants. However, the aim of the grass growth model presented (Johnson and Thornley, 1985) is not to develop a complex soil nitrogen model, therefore a simpler dynamic approach has been adopted (equation A.44).

The Nitrogen dynamic sub-model does not make any distinction between different nitrogen chemical compositions as nitrate, ammonium sulphate. In addition it considers only inorganic nitrogen, which is due to two main assumptions: firstly the main source of organic nitrogen is animals and plants residues which however are broken down into inorganic nitrogen. Secondly over a short period of time, i.e. one growing season, if animals are not present, there should not be a considerable variation in the source of inorganic nitrogen from organic nitrogen in the soil. The rate of contribution of inorganic nitrogen from organic nitrogen breakdown (B_N) is assumed to be temperature dependent (equation A.46).

The application of fertiliser can be a major source of nitrogen to the system, the equation A.45 has been developed in order to account for fertilisation practices and it assumes that the nitrogen is instantaneously and homogeneously distributed through the considered soil volume.

$$\frac{dN_s}{dt} = N_A \delta\left(t - t_F\right) + B_N - \frac{U_N}{d_r \rho_s} - \beta N_s \qquad \text{(A.44)}$$

$$N_A = \frac{A_N}{d_r \rho_s} \qquad \begin{array}{l}\text{if } t = t_F; \\ \text{otherwise } N_A = 0\end{array} \qquad \text{(A.45)}$$

where,

N_A is the fertiliser application function (kg nitrogen (kg dry soil)$^{-1}$

A_N is the fertiliser applied, kg nitrogen m^{-2}

d_r is the roots depth (0.2 m)

ρ_s is the density of dry soil (10^3 kg m^{-3})

$$B_N = B_{20N} \frac{T - T_0}{20 - T_0} \quad \text{(A.46)}$$

where,

B_{20N} is the rate of supply of inorganic nitrogen from soil organic nitrogen at 20°C, kg nitrogen (kg dry soil)$^{-1}$ d^{-1}

T is the mean daily temperature (°C)

T_0 is the temperature at which crops growth stops, 0 (°C)

A.3 C AND N CYCLE IN TERRESTRIAL ECOSYSTEMS: THE PARENT MODEL

The model that is described in this section is an extension of the MBL-MEL model (Rastetter and Shaver, 1997) and it is referred as Parent model (Stapleton, 2004). These two models aim to simulate the Carbon and Nitrogen cycles in terrestrial ecosystems and they can be used to estimate the effects that changes in CO_2 concentration, temperature and Nitrogen input have on the ecosystem Carbon storage.

There are three main differences between the MBL-MEL and the Parent model: firstly the former is characterised by a yearly time-step while the latter has been parameterised with data taken once or twice per growing season which therefore allows the model to have a monthly resolution.

Secondly, the parent model implements a simplified approach to describe the vegetation surface that is active in the uptake of C and N, as the morphological variables SC and SN, used in the MBL-MEL model, have been excluded, since it is difficult to quantify them for an ecosystem which is not dominated by a single specie, as in Svalbard where the samples used for the Parent model have been taken (Stapleton, 2004).

The final aspect of the Parent model, which differs from the original Rastetter and Shaver (1997) model, regards the complexity of the approach adopted to simulate the microbial respiration and the Nitrogen mineralization/immobilisation dynamics. The original approach was unstable during the model parameterisation and therefore unrealistic parameters values were selected in order to reproduce the

nitrogen dynamics (Stapleton, 2005). Therefore the Parent model has been simplified in order to obviate this problem.

The parent model can be divided into three sub-models (Figure A.4):

- *Carbon sub-model*. The carbon absorbed by the vegetation is estimated as function of the CO_2 available and the senescence rate.
- *Nitrogen sub-model*. The estimated inorganic Nitrogen available for plant uptake is function of the plant uptake rate and mineralization/immobilisation due to microbial activity. The microbial biomass is considered to be part of the soil detritus.
- *Acclimation sub-model* allows the model to include the process of acclimation which takes place when the there is a change in the C:N ratio, due to plant development or variations in the ecosystem conditions, i.e. increases of atmospheric CO_2 concentration.

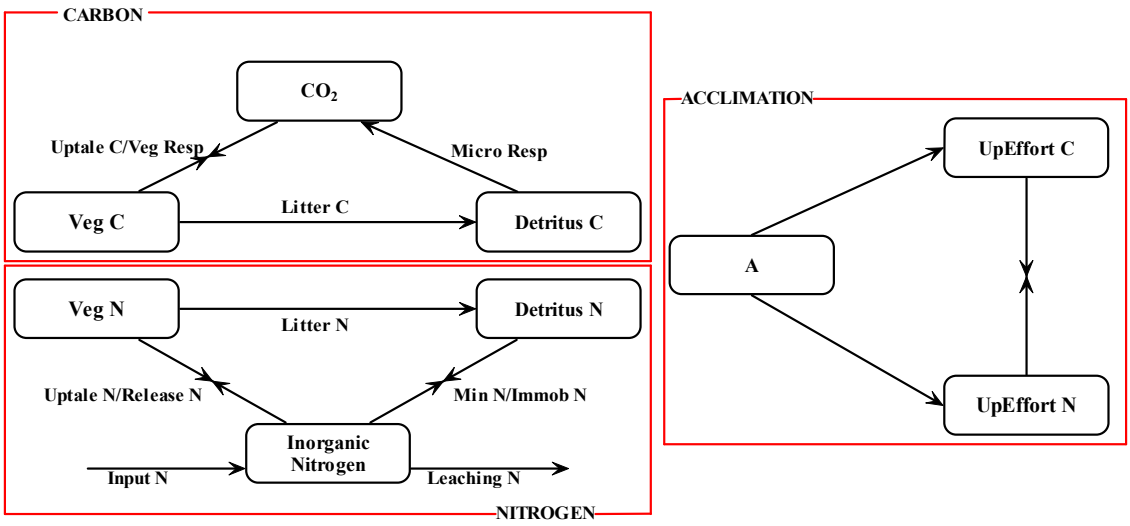


Figure A.23 Three main sub-models form the Parent model: Carbon, Nitrogen and Acclimation sub-model.

A.3.1 Carbon sub-model

The Carbon present in the plant (*VegC*; equation A.47) is assumed to be function of the uptake rate (*UptakeC*; equation A.48), organic and inorganic loss due to senescence (*LitterC*; equation A.49 and plant respiration (*VegResp*; equation A.50), which release carbon into the atmosphere.

$$\frac{dVegC}{dt} = UptakeC - LitterC - VegResp \quad \text{(A.47)}$$

where,

UptakeC is the uptake of inorganic Carbon by the vegetation. It is defined by equation A.48

LitterC refers to the litter loss of organic C from vegetation (gCm^{-2} mo^{-1})

VegResp is the inorganic Carbon release from vegetation (gCm^{-2} mo^{-1})

$$UptakeC = \frac{gC \cdot UpEffortC \cdot CO_2}{kC + CO_2} \quad \text{(A.48)}$$

where,

gC is the plant uptake parameter for Carbon, 76.03 gCm^{-2} mo^{-1}

UpEffortC is an acclimation mechanism, 0.5 (dimensionless)

CO_2 refers to the atmospheric CO_2 concentration, which is a model input.

kC is the Carbon half saturated constant, 475 gCm^{-2} mo^{-1}

$$LitterC = mC \cdot VegC \quad \text{(A.49)}$$

$$VegResp = rC \cdot VegC \quad \text{(A.50)}$$

where,

mC and *rC* are the rate coefficient respectively for litter loss and the vegetation respiration (mo^{-1})

The carbon lost by the plant through senescence, accumulates in the humus layer present on the soil surface. Successively, it is recycled by the microbial population (equation A.52), which through the process of respiration releases it into the atmosphere.

$$\frac{dDetritusC}{dt} = LitterC - MicroResp \qquad \text{(A.51)}$$

where,

MicroResp is the microbial consumption of Carbon (gCm^{-2} mo^{-1}) equation A.52

$$Microresp = MC\left(1 - \Lambda C\right) \qquad \text{(A.52)}$$

where,

MC is the microbial consumption of Carbon (gCm^{-2} mo^{-1}) and it considered as the product between Ψ, which is an optimised parameter with a value of 0.03, and *DetritusC*.

ΛC is the microbial retention efficiency for Carbon (equation A.53; mo^{-1})

UpEffortC is the uptake effort allocated by the plant for the carbon and it is part of the acclimation mechanism.

$$\Lambda C = \frac{\varepsilon C \cdot \theta \cdot MC}{MC + \left(\theta \cdot MN\right)} \qquad \text{(A.53)}$$

where,

εC is the maximum microbial retention for the carbon and it is an optimised parameter, value 0.6

θ is the microbial consumption ratio. It is an optimised parameter with a value of 16.99.

MC and *MN* are optimised parameters which refer to the microbial consumption of carbon and nitrogen and the value are 0.13 and 0.19 respectively ($g\ m^{-2}\ mo^{-1}$).

A.3.2 Nitrogen sub-model

The Nitrogen sub-model simulates the Nitrogen movement within three main pools, described by first order differential equation: the inorganic Nitrogen in the soil (*N*, equation A.54), the Nitrogen in vegetation and finally the Nitrogen that it lost from the plant (*VegN*, equation A.55), due to senescence and it is located in the top part of the soil profile (*DetritusN*, equation A.56).

The inorganic Nitrogen (*N,* g m^{-2}; equation A.54) present in soil has been considered to be replenished by two processes: fertilisation (*InputN*; $kgNha^{-1}$), which is an model input and the Nitrogen added into the soil by gross mineralization (*MinN*; g m^{-2} mo^{-1}) due to microbial activity. On the other hand inorganic Nitrogen is depleted through leaching (*LgN*, g m^{-2} mo^{-1}), immobilisation due to gross microbial uptake (*ImmobN*; g m^{-2} mo^{-1}) and finally vegetation uptake (*UptakeN*; g m^{-2} mo^{-1})

$$\frac{dN}{dt} = InputN + MinN + ReleaseN - LgN - ImmobN - UptakeN \quad \text{(A.54)}$$

The approach used to estimate the vegetation Nitrogen (*VegN*; g m^{-2} equation A.55) and the Nitrogen present in the soil humus layer (*DetritusN*; g m^{-2} equation A.56) resembles the method described previously for Carbon.

$$\frac{dVegN}{dt} = UptakeN - LitterN - ReleaseN \quad \text{(A.55)}$$

$$\frac{dDetritusN}{dt} = LitterN + ImmobN - MinN \quad \text{(A.56)}$$

where,

UptakeN is the uptake of inorganic Nitrogen by the vegetation (gN m^{-2} mo^{-1}). It is defined by equation A.57.

LitterN refers to the loss of Nitrogen from vegetation due to senescence (gN m^{-2} mo^{-1})

ReleaseN is the inorganic Nitrogen release from vegetation (gN m^{-2} mo^{-1})

ImmobN is the Nitrogen immobilised due to gross microbial uptake (gN m^{-2} mo^{-1})

MinN is the microbial Nitrogen consumption (equation A.58 gN m^{-2} mo^{-1})

$$UptakeN = \frac{gN \cdot UpEffortN \cdot N}{kN + N} \quad \text{(A.57)}$$

where,

gN is an optimised parameter referring to the plant uptake, 5.82 gNm^{-2} mo^{-1}

UpEffortN is the effort allocated by the plant for inorganic Nitrogen uptake; it is part of the acclimation mechanism.

N refers to the gross inorganic Nitrogen present in the soil (g m^{-2})

kN is an optimised parameter which refers to the Nitrogen half saturated constant, 0.12 gNm^{-2} mo^{-1}

$$MinN = MN\left(1 - \Lambda N\right) \quad \text{(A.58)}$$

In equation A.58, *MinN* is considered positively correlated to *MN* is the microbial consumption of Nitrogen (gN m^{-2} mo^{-1}) and it considered as the product between Ψ, which is an optimised parameter with a value of 0.03, and *DetritusN*. ΛN is the microbial retention efficiency for Nitrogen (equation A.59; mo^{-1})

$$\Lambda N = \frac{\varepsilon N \cdot MC}{MC + \left(\theta \cdot MN\right)} \quad \text{(A.59)}$$

where,

εN is an optimised parameter which refers to the maximum microbial retention for Nitrogen, value 0.9

θ is the microbial consumption ratio. Optimised value of 16.99.

MC and *MN* are optimised parameters that are the microbial consumption of Carbon and Nitrogen and the value are 0.13 and 0.19 respectively (g m^{-2} mo^{-1}).

A.3.3 Acclimation sub-model

The acclimation process can be defined as the compensatory redistribution of uptake effort in response to a deviation from the optimum element ratios in biomass (Rastetter E.B., 1997). Therefore it allows the vegetation to maintain a nutritional balance when faced with changes in resources availability. This process includes not only physiological and morphological changes in the plant but even genetic and population adaptation may occur. Nevertheless, the Parent model implements a simplified approach assuming that vegetation continuously reallocates uptake effort (*UpEffortC* and *UpEffortN* equatin A.60 and A.61) to maintain an optimum ratio of C:N in the plant biomass.

The optimum C:N ratio (q; equation A.65 is considered to change asymptotically toward the C:N ratio of wood, which is higher than plant tissues, as the biomass increases.

$$\frac{dUpEffortC}{dt} = -a \cdot A \cdot UpEffort^{*} \tag{A.60}$$

$$UpEffort^{*} = UpEffortC \qquad \text{if } A > 0$$
$$UpEffort^{*} = UpEffortN \qquad \text{if } A < 0$$

$$\frac{dUpEffortN}{dt} = -\frac{dUpEffortC}{dt} \tag{A.61}$$

where,

$$A = \ln \frac{VegC}{q \cdot VegN} + \tau \left(GrC - GrN \right) \tag{A.62}$$

$$GrC = -\frac{dVegC}{dt} \cdot \frac{1}{VegC} \tag{A.63}$$

$$GrN = -\frac{dVegN}{dt} \cdot \frac{1}{VegN} \tag{A.64}$$

$$q = \frac{qstems \cdot VegC}{kq + VegC} \tag{A.65}$$

A.5 C AND N CYCLE IN TERRESTRIAL ECOSYSTEMS: THE TEMP MODEL

The Temp model, developed by Stapleton (2005), is an extension of the Parent model and it has been designed to account for the temperature dependency of plant physiological processes and consequently increase the mechanistic credibility of the Parent model.

Temperature response functions (TRFs; equation A.66) have been incorporated in the Parent model in order to make *UptakeC*, *MicroResp*, *UptakeN*, *ReleaseN*, *MinN*, *VegResp* and *ImmobN* temperature dependent.

$$Qx(T) = e^{qx(T - T_O x)} \left(\frac{T_M x - T}{T_M x - T_O x} \right)^{qx(T_M x - T_O x)} \tag{A.66}$$

where,

qx is the curvature parameter (°C^{-1})

T is the mean temperature (°C)

$T_M x$ is the maximum temperature (°C)

$T_O x$ is the optimum temperature (°C)

x is *UptakeC*, *MicroResp*, *UptakeN*, *ReleaseN*, *MinN*, *VegResp* and *ImmobN*

BIBLIOGRAPHY

ABSALOM, J. P. CROUT, N. M. J. and YOUNG, S. D. (1996) Modeling radiocesium fixation in upland organic soils of northwest England, *Environmental Science & Technology* 30, 2735-2741

ABSALOM, J. P. YOUNG, S. D. CROUT, N. M. J. SANCHEZ, A. WRIGHT, S. M. SMOLDERS, E. NISBET, A. F. and GILLETT, A. G. (2001) Predicting the transfer of radiocaesium from organic soils to plants using soil characteristics, *Journal of Environmental Radioactivity* 52: 31-43.

ABSALOM, J.P. YOUNG, S. D. CROUT, N. M. J. NISBET, A. F. WOODMAN, R. F. M. SMOLDERS, E. and GILLETT, A. G. (1999) Predicting soil to plant transfer of radiocesium using soil characteristics. *Environmental Science & Technology* 33, 1218-1223

ABSOLOM, J. P., YOUNG, S. D., CROUT, N. M. J., SANCHEZ, A., WRIGHT, S. M., SMOLDERS, E., NISBET, A. F. and GILLETT, A. G. (2001) Predicting the transfer of radiocaesium from organic soils to plants using soil characteristics, *Journal of Environmental Radioactivity* 52: 31-43.

AHMAN, B., WRIGHT, S.M. and HOWARD, B.J. (2001) Effect of origin of radiocaesium on the transfer from fallout to reindeer meat, *The science of the total environment* 278: 171-81.

ALBERS, B. P. STEINDL, H. SCHIMMACK, W. and BUNZL, K. (2000) Soil-to-plant and plant-to-cows milk transfer of radiocaesium in alpine pastures: significance e of seasonal variability. *Chemosphere* 41, 717-723

APSIMON, H. M. WILSON, J. J. N. and SIMMS, K. L. (1989) Analysis of the dispersal and deposition of radionuclides from Chernobyl across Europe. Proceedings of the Royal Society. London A 425, 365-405

AVILA, R., BERGMAN, R., SCIMONE, M., FESENKO, S., SANCHAROVA, N. and MOBERG, L. (2001) A comparison of three models of 137Cs transfer in forest ecosystems, *Journal of Environmental Radioactivity* 55/3: 315-27.

BAEZA, A. GUILLÉN, F. J. SALAS, A. and MANJÓN, J. L. (2005) Distribution of radionuclides in different parts of a mushroom: Influence of the degree of maturity.

Science of The Total Environment In Press, Corrected Proof, Available online 8 September 2005

BARBER, S.A. (1984) Soil nutrient bioavailability. A mechanistic approach. John Wiley & Sons: NY

BARNES, E. C. (2000) Ockham's Razor and the Anti-Superfluity Principle, *Erkenntnis* 53/3: 353 - 74.

BASGALL M. (2005) http://www.sciencedaily.com/releases/2005/10/051001100950.htm (last accessed, October 2005)

BELLI, M. GERZABECK, M. WIRTH, E. BUNZL, K. SHAW, G. RAFFERTY, B. and DELVAUX, B. (1999). *SEMINAT, Long-term dynamics of radionuclides in semi-natural environments: Derivation of parameters and modelling*., SEMINAT final report. ANPA, Italy: European Commission – Nuclear Fission Safety Programme contract n° FI4P-CT95-0022.

BOZDOGAN, H. (2000) Akaike's Information Criterion and Recent Developments in Information Complexity. *Journal of Mathematical Psychology* 44, 62-91

BRIDGES, E. M. (1997) *World soils*. 3rd edition, Cambridge University Press, Cambridge.

BROOKS, R. J. and TOBIAS, A. M. (1996) Choosing the best model: Level of detail, complexity, and model performance, *Mathematical and Computer Modelling* 24/4: 1-14.

BRUGNACH, M., BOLTE, J. and BRADSHAW, G. A. (2003) Determining the significance of threshold values uncertainty in rule-based classification models, *Ecological Modelling* 160/1-2: 63-76

COLACINO M. (2005) http://150.146.47.106/rivistaonline/base_esperto.asp?IDRubrica=6&nomefile=01_15_2005 (last accessed, October 2005)

COX, G. M. (2004) *Managing radioactively contaminated land: a method to assist the design of long-term remediation strategies*. PhD thesis, University of Nottingham.

CROSETTO, M. TARANTOLA, S. and SALTELLI, A. (2000) Sensitivity and uncertainty analysis in spatial modelling based on GIS. *Agriculture, Ecosystems & Environment* 81, 71-79

CRYSTAL BALL (1996) Crystal Ball 2000 professional, version 5.0 Users Manual. Decisioneering Inc., Aurora, CO.

DAVISON, W., SPEZZANO, P. and HILTON, J. (1993) Remobilization of caesium from freshwater sediments, *Journal of Environmental Radioactivity* 19: 109-24.

DeCORT, M. (1998) **http://rem.jrc.cec.eu.int/downloads/Atlas/** (last accessed October 2005)

EC, (1995) **http://eusoils.jrc.it/** (last accessed October 2005)

EHLKEN, S and KIRCHNER, G (2002) Environmental processes affecting plant root uptake of radioactive trace elements and variability of transfer factor data: a review, *Journal of Environmental Radioactivity* 58: 97-112.

ERHARDT, J. PASLER-SAUER, J. SCHULE, O. BENZ, G. RAFAT, M. and RICHTER, J. (1993) Development of RODOS, a comprehensive decision support system for nuclear emergencies in Europe – an overview. *Radiation Protection Dosimetry* 50, 195-203

FACCHINELLI, A., GALLINI, L., BARBERIS, E., MAGNONI, M. AND HURSTHOUSE, A.S. (2001) The influence of clay mineralogy on the mobility of radiocaesium in upland soils of NW Italy, *Journal of Environmental Radioactivity* 56: 299-307.

FAO-AGLL (2000) **http://www.fao.org/ag/agl/agll/infotech.htm** (last accessed October 2005)

FLETCHER R. (1987) Practical Methods of optimisation, 2nd edition, John Wiley & Sons: Chichester, NY, Brisbane.

FORSTER, M. R. (2000) Key Concepts in Model Selection: Performance and Generalizability. *Journal of Mathematical Psychology* 44, 205-231

FREY, H. C. and Patil, S. R. (2002) Identification and review of sensitivity analysis methods *Risk Analysis*. 22, 553-577

FREUND, R. J. and WILSON, W. J. (2003) *Statistical methods*. 2nd edition, Accademi Press: London

FRISSEL, M. J., DEB, D. L., FATHONY, M., LIN, Y. M., MOLLAH, A. S., NGO, N. T., OTHMAN, I., ROBISON, W. L., SKARLOU-ALEXIOU, V., TOPCUOLU, S., TWINING, J. R., UCHIDA, S. and WASSERMAN, M. A. (2002) Generic values for soil-to-plant transfer factors of radiocesium, *Journal of Environmental Radioactivity58* 58/2-3: 113-28.

GELMAN, A., CARLIN, J.B., STERN, H.S. and RUBIN, D.B. (2004) Bayesian data analysis. 2nd ed. Chapman & Hall: London.

GEORGE, L. (1977) Tests for system complexity, *International Journal of general systems* 3: 227-32

GERSHENFELD, N. (1999) The nature of mathematical modelling. Cambridge University Press, Cambridge

GERZABEK, M. H. STREBL, F. EHLKEN, S. and KIRCHNER, G. (2003) *Radiation in the soil-plant system*. In *Handbook of processes and modeling in the soil-plant system*. pp 149-175 (Eds. D. K. Benbi, R. Nieder) Food productions Press: NY, London

GIBSON, G. J., KLECZKOWSKI, A. and GILLIGAN, C. A. (2004) Bayesian analysis of botanical epidemics using stochastic compartmental models, *Proc. Nat. Acad. Sciences*, 101: 12120-12124.

GOMMERS, A., GÄFVERT, T., SMOLDERS, E., MERCKX, R. and VANDENHOVE, H. (2005) Radiocaesium soil-to-wood transfer in commercial willow short rotation coppice on contaminated farm land, *Journal of Environmental Radioactivity,* 78: 267-87.

GOOR, F. and AVILA, R. (2003) Quantitative comparison of models of 137Cs cycling in forest ecosystems, *Environmental Modelling & Software* 18/3: 273-79.

GRUNWALD, P. (2000) Model Selection Based on Minimum Description Length. *Journal of Mathematical Psychology* 44, 133-152

HELLING, C.S. CHESTERS, G. and COREY, R.B., (1964) Contribution of organic matter and clay to soil cation exchange capacity as affected by the pH of the saturating solution. *Soil Science Society America* 28, 517–520.

HELTON, J. C. JOHNSON, J. D. ROLLSTIN, J. A. SHIVER, A. W. and SPRUNG, J. L. (1995) Uncertainty and sensitivity analysis of chronic exposure results with the MACCS reactor accident consequence model. *Reliability Engineering & System Safety* 50, 137-177

HELTON, J.C. DAVIS F.J. and JOHNSON J.D. (2005) A comparison of uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling. *Reliability Engineering & System Safety* 89, 305-330

HOWARD, B. SMOLDERS, E. GIL, J. M. VOIGT, G. STRAND, P. and CROUT, N. (1999) *Spatial analysis of vulnerable ecosystem in Europe: Spatial and dynamic prediction of radiocaesium fluxes into European foods (SAVE)*. Final report for the SAVE project. UK. European Commission – Nuclear Fission Safety Programme contract n° F14-PCT95-001

IAEA. INTERNATIONAL ATOMIC ENERGY AGENCY (1994) *Handbook of parameters values for the prediction of radionuclide transfer in temperate environments*. Technical report series No. 364. IAEA, Vienna, Austria

IAEA. INTERNATIONAL ATOMIC ENERGY AGENCY (1996) *Validation of models using Chernobyl fallout data from southern Finland: Scenario S*. IAEA-TECDOC-904. IAEA, Vienna, Austria

IAEA. INTERNATIONAL ATOMIC ENERGY AGENCY (2002) *BIOMASS - Modelling the migration and accumulation of radionuclides in forest ecosystems* IAEA-BIOMASS-1. IAEA, Vienna, Austria

JOHNSON, I.R. and THORNLEY, H.M. (1983) Vegetative crop growth model incorporating leaf area expansion and senescence and applied to grass, *Plant, Cell and Environment* 6: 721-29.

JOHNSON, I.R. and THORNLEY, H.M. (1984) A model of instantaneous and daily canapy photosynthesis, *Journal of Theoretical Biology* 107: 531-45.

JOHNSON, I.R. and THORNLEY, H.M. (1985) Dynamic model of the response of a vegetative grass crop to light, temperature and nitrogen. *Plant, Cell and Environment* 8: 485-99

JONES P. N. and CARBERRY P. S. (1994) A technique to develop and validate simulation models. *Agricultural Systems* 46, 427-442

KAMMERER, L. HIERSCHE, L. and WIRTH, E. (1994) Uptake of radiocaesium by different species of mushrooms. *Journal of Environmental Radioactivity* 23, 135-150

KAMMERER, L., HIERSCHE, L. and WIRTH, E. (1994) Uptake of radiocaesium by different species of mushrooms, *Journal of Environmental Radioactivity* 23: 135-50.

KARLÉN, G. JOHANSON, K. J. and BERGSTRÖM, R (1991) Seasonal variation in the activity concentration of 137Cs in Swedish Roe-deer and in their daily intake, *Journal of Environmental Radioactivity* 14/2: 91-103.

KENDALL, M. and GIBBONS, J.D. (1990) *Rank Correlation Methods* (London, Melbourne, Auckland: Hodder & Stoughton).

KIEFER, P. PRÖHL, G. MÜLLER, H. LINDNER, G. DRISSNER, J. and ZIBOLD, G. (1996) Factors affecting the transfer of radiocaesium from soil to roe deer in forest ecosystems of southern Germany. *Science of The Total Environment*. 192, 49-61

KUCZERA, G. and PARENT, E. (1998) Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology* 211, 69-85

LINDNER, G., DRISSNER, J., HERRMANN, T., HUND, M., ZECH, W., ZIBOLD, G. and ZIMMERER, R. (1994) Seasonal and regional variations in the transfer of cesium radionuclides from soil to roe deer and plants in a prealpine forest, *Science of The Total Environment* 157/1-3: 189-96.

LINKOV, I. YOSHIDA, S. and STEINER, M. (2000) Fungi contaminated by radionuclides : critical review of approaches to modelling.
www.irpa.net/irpa10/cdrom/00967.pdf (last accessed October 2005)

MATTHEWS, R. METHLEY, J. ALEXANDER, M. JOKIEL, P. and SALISBURY, I. (1996) Site classification and yield prediction for lowland sites in England and Wales. Forestry Commission and Ministry of Agriculture, Fisheries and Food, CSA 2119, Surrey, UK

McCARTHY, P. (2005) http://news.independent.co.uk/uk/environment/article15380.ece (last accessed, October 2005)

McKAY, M.D. (1992) Latin Hypercube sampling as a tool in uncertainty analysis of computer models, in J.J. Swain, D. Goldsman, R.C. Crain and J.R. Wilson (eds), *Winter Simulation Conference*.

MELIN, J., WALLBERG, L. and SUOMELA, J. (1994) Distribution and retention of cesium and strontium in Swedish boreal forest ecosystems, *The Science of The Total Environment,* 157: 93-105.

MITCHELL, M. (1995) Genetic algorithms: An overview, *Complexity* 1, 31-39.

MONTE, L. (1994) Evaluation of radionuclide transfer functions from drainage basins of fresh water systems, *Journal of Environmental Radioactivity* 26: 71-82.

MONTE, L. BRITTAIN, J. E. HAKANSON, L. and DIAZ, E. G. (1999) *MOIRA: models and methodologies for assessing the effectiveness of countermeasure in complex aquatic system contaminated by radionuclides*. RT/AMB/99/1. ENEA, Rome, Italy

MONTE, L., BRITTAIN, J. E., HÅKANSON, L., HELING, R., SMITH, J. T. and ZHELEZNYAK, M. (2003) Review and assessment of models used to predict the fate of radionuclides in lakes, *Journal of Environmental Radioactivity* 69/3: 177-205.

MONTE, L., HÅKANSON, L., BERGSTRÖM, U., BRITTAIN, J. and HELING, R. (1996) Uncertainty analysis and validation of environmental models: the empirically based uncertainty analysis, *Ecological Modelling* 91/1-3: 139-52.

MONTERO, M. MORALEDA, M. CLAVER, F. VASQUEZ, C. and GUTIERREZ, J. (2001) *Methodology for decision making in environmental restoration after nuclear accidents: TEMAS system (2.1)*. CIEMAT, Madrid, Spain.

MOSCHANDREAS, D. J. and KARUCHIT, S. (2002) Scenario–model–parameter: a new method of cumulative risk uncertainty analysis, *Environment International* 28/4: 247-61.

MULLER, H and PROHL, G. (1993) ECOSYS-87 - A dynamic-model for assessing radiological consequences of nuclear accidents. *Health Physics* 64, 232-252

MYUNG, I. J. (2000) The Importance of Complexity in Model Selection, *Journal of Mathematical Psychology* 44: 190-204.

NIMIS, P. L. (1996) Radiocaesium in plants of forest ecosystems. *Studia Geobotanica* 15, 3-49

NISBET, A. F. and WOODMAN, R. F. M. (2000) Soil-to-plant transfer factors for radiocesium and radiostrontium in agricultural systems. *Health Physics* 78, 279-288

NISBET, A. F. WOODMAN, R. F. M. and HAYLOCK, R. G. E. (1999) *Recommended soil-to-plant transfer factors for radiocaesium and radiostrontium for use in arable systems*. NRPB-R304. National Radiological Protection Board, Didcot, UK

OBERKAMPF, W. L., DELAND, S. M., RUTHERFORD, B. M., DIEGERT, K. V. and ALVIN, K. F. (2002) Error and uncertainty in modeling and simulation, *Reliability Engineering & System Safety* 75/3: 333-57.

PAPADOPOULOS, E. C. and YEUNG, H. (2001) Uncertainty estimation and Monte Carlo simulation method, *Flow Measurement and Instrumentation* 12/4: 291-98.

PITT, M. A. and MYUNG, I. J. (2002) When a good fit can be bad, *Trends in Cognitive Sciences* 6/10: 421-25.

PORTE, A. and BARTELINK, H. H. (2002) Modelling mixed forest growth: a review of models for forest management. *Ecological Modelling*. 150, 141-188

RASTETTER, E. B. AGREN, G. I. and SHAVER, G. R. (1997) Responses of N-limited ecosystems to increased CO_2: a balanced-nutrition, coupled-element-cycles model. *Ecological Society of America* 7, 44-460

REID T.D. and CROUT N.M.J. (2004) Selecting Appropriate Models of Antarctic Lake Ice Dynamics. From Proceedings of the 2004 *Environmental Modelling and Simulation* conference

SALTELLI, A. CHAN, K. and SCOTT, E. M. (2000a) *Sensitivity Analysis*. John Wiley and Sons ltd: Chichester.

SALTELLI, A., TARANTOLA, S. and CAMPOLONGO, F. (2000b) Sensitivity Analysis as an ingredient of modeling, *Statistical Science* 15/4: 377-395.

SANCHEZ, A. L. SMOLDERS, E. VAN DEN BRANDE, K. MERCKX, R. WRIGHT S. M. and NAYLOR, C. (2002) Predictions of in situ solid/liquid distribution of radiocaesium in soils. *Journal of Environmental Radioactivity*. 63, 35-47

SANCHEZ, A. L., WRIGHT, S. M., SMOLDERS, E., NAYLOR, C., STEVENS, A., KENNEDY, V. H., DODD, B. A., SINGLETON, D. L. and BARNETT, C. L. (1999) High plant uptake of radiocesium from organic soils due to Cs mobility and low soil K content, *Environmental Science and Technology* 33: 2752-57.

SANCHEZ, A. L., WRIGHT, S. M., SMOLDERS, E., NAYLOR, C., STEVENS, A., KENNEDY, V. H., DODD, B. A., SINGLETON, D. L. AND BARNETT, C. L. (1999) High plant uptake of radiocesium from organic soils due to Cs mobility and low soil K content, *Environmental Science and Technology* 33: 2752-57.

SCHRUBEN, L. and YUCESAN, E. (1993) Complexity of simulation models: a graph theoretic approach. From Proceedings of the 1994 *Winter Simulation* Conference

SIPPOLA, J. and YLI-HALLA, M. (1998) *Status of soil mapping in Finland*. European soil Bureau, Jokioinen, Finland

SKUTERUD, L., GAARE, E., EIKELMANN, I. M., HOVE, K. and STEINNES, E. (2005) Chernobyl radioactivity persists in reindeer, *Journal of Environmental Radioactivity* 83: 231-52.

SKUTERUD, L., TRAVNIKOVA, I.G., BALONOV, M.I., STRAND, P. and HOWARD, B.J. (1997) Contribution of fungi to radiocaesium intake by rural populations in Russia, *Science of The Total Environment* 193: 237-42.

SHAW, G. AVILA, R. FESENKO, S. DVORNIK, A. and ZHUCHENKO, T. (2002) Modelling the Behaviour of Radiocaesium in Forest Ecosystems. In *Modelling Radioactivity in the Environment*, ed. E. M. Scott, Elsevier Science Publishers.

SIVIA, D.S. (1996) Data Analysis. A Bayesian Tutorial. Oxford Science Publications: Oxford.

SMITH, J. T. and BERESFORD N. A. (2005) *Chernobyl Catastrophe and consequences*, Springer: Chichester, UK

SMITH, J. T. BERESFORD, N. A. SHAW, G. G. and MOBERG, L. (2005a) Radioactivity in terrestrial ecosystems. In *Chernobyl Catastrophe and consequences*. Ed Smith J. and Beresford N. A. Springer: Chichester, UK

SMITH, J. T. VOITSEKHOVITCH, O. V. KONOPLEV, A. V. T. and KUDELSKY, A. V. (2005b) Radioactivity in aquatic systems. In *Chernobyl Catastrophe and consequences*. ed Smith J. and Beresford N. A. Springer: Chichester, UK

SMITH, J. T., COMANS, R. N. J., IRELAND, D. G., NOLAN, L. T. and HILTON, J. (2000) Experimental and in situ study of radiocaesium transfer across the sediment–water interface and mobility in lake sediments, *Applied Geochemistry* 15: 833-48.

SMOLDERS, E., VANDENBRANDE, K. and MERCKX, R. (1997) Concentrations of Cs-137 and K in soil solution predict the plant availability of Cs-137 in soils, *Environmental Science and Technology* 31/12: 3432-38.

STAPLETON , L. M. (2004) *Modelling Carbon and Nitrogen fluxes for two terrestrial ecosystems on Svalbard*. PhD thesis, University of Nottingham.

STOCKLE, C. O. (1992) Canopy photosynthesis and transpiration estimate using radiation interception models with different levels of detail, *Ecological Modelling* 60/1: 31-44.

SUMMERHAYES, C. (2005) http://comment.independent.co.uk/letters/article296825.ece (last accessed, October 2005)

TURANYI, T. (1990) Sensitivity analysis of complex kinetic systems. Tools and applications. *Journal of Mathematical Chemistry* 5: 203-48.

UNSCEAR, UNITED NATIONS SCIENTIFIC COMMITTEE ON THE EFFECTS OF ATOMIC RADIATION (1988) *Sources, effects and risks of ionizing radiation.* Unite Nations.

UNSCEAR, UNITED NATIONS SCIENTIFIC COMMITTEE ON THE EFFECTS OF ATOMIC RADIATION (1993) *Exposures from man-made sources of radiation.* Unite Nations.

VALCKE, E. and CREMERS, A. (1994) Sorption-desorption dynamics of radiocaesium in organic matter soils. Science of the Total Environment 157, 275-283

VARDOULAKISA, S., FISHER, B. E. A., GONZALEZ-FLESCAA, N. and PERICLEOUSB, K. (2002) Model sensitivity and uncertainty analysis using roadside air quality measurements, *Atmospheric Environment* 36/13: 2121-34.

VILIC, M. BARISIC, D. KRALJEVIC, P. and LULIC, S. (2005) 137Cs concentration in meat of wild boars (Sus scrofa) in Croatia a decade and half after the Chernobyl accident. *Journal of Environmental Radioactivity* 81, 55-62

WEBSTER, D. B. PADGETT, M. L. HINES, G. S. and SIROIS, D. L. (1984) Determining the level of detail in a simulation model. A case study. *Computer and Industrial Engineering*. 8, 215-225

WHICKER, F. W. and KIRCHNER, T. B. (1987) Pathway: a dynamic food-chain model to predict radionuclide ingestion after fallout deposition. *Health Physics*. 52, 717-737

ZHU, Y-G. and SMOLDERS, E. (2000) Plant uptake of radiocaesium: a review of mechanisms, regulation and application, *Journal of Experimental Botany* 51: 1635-45.

ZIBOLD, G., DRISSNER, J., KAMINSKI, S., KLEMT, E. and MILLER, R. (2001) Time-dependence of the radiocaesium contamination of roe deer: measurement and modelling, *Journal of Environmental Radioactivity* 55: 5-27.

ZUCCHINI, W. (2000) An Introduction to Model Selection. *Journal of Mathematical Psychology*. 44, 41-61